Proceedings of the ESSLLI 2014 Student Session

26th European Summer School in Logic, Language & Information

August 11–22, 2014, Tübingen, Germany

Ronald de Haan (editor)

Preface

These proceedings contain the papers presented at the Student Session of the 26th European Summer School in Logic, Language and Information (ESSLLI 2014), taking place at the Eberhard Karls University of Tübingen in August 2014. As the Student Session has been organized this year for the nineteenth time, it has been part of the ESSLLI tradition for almost two decades. It offers an excellent venue for students to present their work on a diverse range of topics at the crossroads of logic, language and information. This is attested by the large number of high quality submissions. We received 67 submissions, 47 of which were submitted for oral presentation, and 20 of which were submitted for a poster presentation. At the Student Session, 16 of these submissions were presented orally and 6 submissions were presented in the form of a poster.

I would like to thank each of the co-chairs, as well as the area experts, for all their invaluable help in the reviewing process and organization of the Student Session. Without them, the Student Session would not have been able to take place. I would also like to thank the ESSLLI Organizing Committee for organizing the entire summer school, and catering to all our needs. They have been wonderful hosts to the Student Session. Thanks go to the chair of the previous editions of the Student Session as well, in particular to Margot Colinet, for their advice. As in previous years, Springer-Verlag has generously offered prizes for *Best Paper* and *Best Poster Awards*, and for this we are very grateful. Most importantly, thanks to all those who submitted papers, for they are the ones that make the Student Session the exciting event and great breeding ground that it is.

August 2014

Ronald de Haan
Chair of the ESSLLI 2014 Student Session

Organization

Program Committee

Chair:

Ronald de Haan	Technische Universität Wien

Co-chairs (Logic & Computation):

Zoé Christoff	Universiteit van Amsterdam
Aybüke Özgün	Université de Lorraine

Co-chairs (Logic & Language):

Philip Schulz	Universiteit van Amsterdam
Thomas Brochhagen	Universität Düsseldorf, Universiteit van Amsterdam

Co-chairs (Language & Computation):

Miriam Kaeshammer	Universität Düsseldorf
Ramon Ziai	Universität Tübingen

Area experts

Co-chairs (Logic & Computation):

Jakub Szymanik	Universiteit van Amsterdam
Balder ten Cate	University of California, Santa Cruz

Co-chairs (Logic & Language):

Michael Franke	Universität Tübingen

Co-chairs (Language & Computation):

Roger Levy	University of California, San Diego

Contents

Logic & Computation

Logic & Language

Language & Computation

⋆ poster presentations

Connecting the Categorical and the Modal Logic Approaches to Quantum Mechanics

Giovanni Cinà

Institute for Logic, Language and Computation
University of Amsterdam

Abstract This paper aims at connecting the two research programs known as Categorical Quantum Mechanics and Dynamic Quantum Logic. This goal is achieved in three steps. First we define a procedure to extract a Modal Logic frame from a small category and a functor into the category of sets and relations. Second, we extend such methodology to locally small categories. Third, we apply it to the category of finite-dimensional Hilbert spaces to recover the semantics of Dynamic Quantum Logic.
This procedure prompts new lines of research. In the case of Hilbert spaces, we investigate how to obtain richer semantics, containing probabilistic information. We design a logic for this semantics and prove that, via translation, it preserves the validities of Dynamic Quantum Logic.

1 Introduction

The development of Quantum Computation and Information has caused a new wave of studies in Quantum Mechanics: the possibility of defining quantum algorithms, and the fact that some of them outperform their classical counterparts, has elicited new theoretical questions. In particular, we are interested in crafting a formalism that captures the features of quantum processes. The intended tool that we would want to obtain from such a formalism is a formal system capable of proving the correctness of quantum algorithms.

We present here two frameworks that have been proposed for such a task: Categorical Quantum Mechanics and Dynamic Quantum Logic. They share the same theoretical aim and the same intended application; this constitutes a natural motivation to investigate the connection between the two.

The first research program, pioneered by Abramsky and Coecke, is a study of Quantum Mechanics through the lenses of Category Theory. Their work started from the analysis of the categorical structure of the category of finite-dimensional Hilbert spaces and linear maps.[1] In the last decade this research project has produced many results and a renewed interest in symmetric monoidal categories, the categories used to model compound systems.

[1] The first paper on Categorical Quantum Mechanics is [1]. See [2] for an extensive survey.

The second approach, proposed by Baltag and Smets, exploits the formalism of PDL to represent quantum algorithms and to design a proof system able to prove their correctness. This approach is connected with both the traditional logical studies of the foundations of Quantum Mechanics, the so-called standard Quantum Logic, and the "Dynamic Turn" in Logic, that is, the use of modal logics to reason about processes and information. This research group has proposed different logics for this task; here we focus on LQP, the Logic of Quantum Programs, and its compound version LQP^n.[2] The semantics of these logics are relational structures called Dynamic Quantum Frames, namely relational versions of Hilbert spaces.

In order to relate these two research programs, we show how the frames of LQP and LQP^n are related to the categories studied by Abramsky and Coecke. First, we define a procedure to extract a Modal Logic frame from a small category and a functor into the category **Rel** of sets and relations. Second, we extend such methodology to locally small categories. Third, we apply it to the category of finite-dimensional Hilbert spaces to recover the semantics of LQP and LQP^n. The bridge between the category of finite-dimensional Hilbert spaces and the semantics of these logics depends on the choice of a specific functor into **Rel**.

This construction suggests new lines of research. In the case of Hilbert spaces, we show how to obtain a richer semantics, containing probabilistic information, with the choice of a different functor. The indended use for such semantics is the formalization of protocols where probabilities play an essential part. We design a language able to capture this additional probabilistic information and show that, via translation, the set of validities of the frames arising from the new functor preserves the validities of LQP and LQP^n. This means that the correctness proofs casted in the language of LQP^n can be trasferred in the new language.

The paper is structured as follows. The first two sections are devoted to an outline of the two research programs, Categorical Quantum Mechanics and Dynamic Quantum Logic. In the third section we outline the methodology to extract a Modal Logic frame from a small category and a functor into **Rel**. In the fourth part we show how to apply our methodology to the category of finite-dimensional Hilbert spaces, describing the formal bridge between the two aforementioned approaches, and we expand on the possibility to define and study a richer semantics containing probabilistic information.

This article is based on [7]; we refer to it for a detailed explanation of the results presented here, further discussion and examples. In what follows we employ concepts and notation from Category Theory, Modal Logic and Quantum Computing; we direct the reader to the textbooks in the references (respectively [8], [6] and [9]) or to the appendix in [7] for the necessary background in these areas.

[2] This line of research is developed in multiple papers, we refer to [5] and [4] in particular.

2 Categorical Quantum Mechanics

In their paper [2], Abramsky and Coecke outline a study of foundations of Quantum Mechanics from a category-theoretic perspective. The target of this study is $\mathbf{FdHil}_{\mathbb{C}}$, the category having as objects finite-dimensional Hilbert spaces over the field of complex numbers and as morphisms linear maps.[3] This category can be thought of as the formal environment where Quantum Computing takes place.[4]

The key observation is the following:

Theorem 1 ([2]). *The category* **FdHil** *is a dagger compact closed category with biproducts.*

This in particular means that **FdHil** is:

(1) a symmetric monoidal category
(2) a compact closed category
(3) a dagger category
(4) a category with biproducts

and furthermore that all these layers of structure coexist together, namely that the category satisfies some coherence conditions (see [2] and also [8], pp. 158-9).

Intuitively, a symmetric monoidal category is a category equipped with an operation to mold objects into compound objects. In **FdHil** this role is fulfilled by the tensor product. Monoidal categories have a special object I which is the unit of the operation, in **FdHil** this is the one-dimensional Hilbert space $\mathbb{C}$. This unit object can be used to characterize scalars in general as morphisms $I \to I$; this definition specializes well, since the linear maps of type $\mathbb{C} \to \mathbb{C}$ correspond indeed to the scalars in $\mathbb{C}$.

A compact closed category is a category having, for each object, a dual object with particular properties. In **FdHil** these are the conjugate spaces, that is, the spaces in which scalars and inner product are the complex conjugate with respect to the original space. Dagger categories have a contravariant, identity-on-objects and involutive endofunctor, namely an operation $\dagger$ that modifies only morphisms and switches domains and codomains. This corresponds to the conjugate-transpose of a linear map in **FdHil**. Via this additional structure we can characterize unitary maps as isomorphisms such that $f^{-1} = f^{\dagger}$ and self-adjoints maps as morphisms such that $f = f^{\dagger}$. This also suggests the abstract characterization of projectors as self-adjoint morphisms such that $f \circ f = f$.

Finally, a category with biproducts is a category with a distinguished object, called *zero* object, and an operation to merge objects together. Contrarily to the monoidal operation, biproducts stand for objects that are completely determined by their components. The zero object in **FdHil** is the 0-dimensional vector space, while the biproduct is the direct sum of Hilbert spaces.

[3] We will drop the subscript in what follows.
[4] The limitation to finite-dimensional Hilbert spaces is a rather strandard one in Quantum Computation, see for example [9].

The central ingredients of Quantum Mechanics can be recovered in this categorical framework, and we can give an abstract representation of quantum protocols. Furthermore, we can prove the correctness of a protocol via the commutation of the appropriate diagram.

2.1 Example: Teleportation Protocol

To exemplify the way in which a quantum protocol is represented categorically, we have a closer look at the Teleportation protocol. The treatment will be partially informal, as we have not presented enough formal background for a thorough explanation.

The Teleportation protocol describes a technique to transfer a quantum state from one agent, called with the fictional name Alice, to another agent, named Bob. This procedure does not require the existence of a quantum communication channel between Alice and Bob, but a classic communication channel is needed. The Hilbert space describing the system is the tensor product of three 2-dimensional systems $H = H_1 \otimes H_2 \otimes H_3$, that is, it is a space consisting of three qubits. We suppose Alice and Bob possess one qubit each of an entangled Bell state $\beta_{00} \in H_2 \otimes H_3$. Alice also has a qubit q_1 given by a state of H_1.

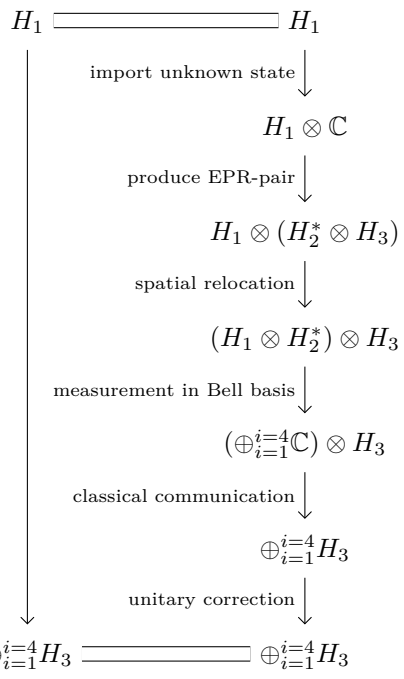


Figure 1: Quantum Teleportation (from [2])

After obtaining their part of the entangled Bell state, Alice and Bob become separated; we assume $H_1 \otimes H_2$ is the part of the system available to Alice and H_3 is the part available to Bob. The goal of Alice is to teleport her additional qubit to the location of Bob, i.e., to turn the state of H_3 into the initial state of H_1.

In order to do so, Alice performs a measurement in the Bell basis, that is, a measurement such that each projector projects into one of the vectors of the Bell basis, on her two qubits. The result of this measurement is a pair of classical bits. The action that Bob has to perform on q_3 to obtain the initial q_1 depends on the measurement outcome obtained by Alice, so using the classical communication channel between them, she sends this pair of classical bits to Bob, who performs a quantum gate according to the following table:

$$00 \longrightarrow Id$$
$$01 \longrightarrow Z$$
$$10 \longrightarrow X$$
$$11 \longrightarrow XZ$$

The final qubit q_3 of Bob is then equal to the initial q_1.

The Teleportation protocol is represented in Figure 1. The arrow on the left of the diagram encodes the expected effect of the protocol, namely the fact that the qubit in H_1 is transferred to the system H_3 in all four possible evolutions of the system (the number 4 is given by the number of possible outputs of the measurement on the Bell basis). The arrow on the right side depicts the protocol itself. The commutation of this diagram expresses the correctness of the algorithm: it amounts to a universal statements on all the vectors of the Hilbert space H_1, proving that the protocol works *for every input vector*, that is, for every state of the system H_1.

The commutation of this diagram can be proved in this framework, see Theorem 30 in [2] p.41.

3 Dynamic Quantum Logic

The Logic of Quantum Programs LQP and it compound version LQP^n were designed by Baltag and Smets to express quantum algorithms and prove their correctness (see [5] and [4]). The core ideas behind this logics are two. First, we can see the states of a physical system as states of a Modal Logic frame. Second, the dynamics of the system can be captured by means of a PDL-style formalism, i.e., a modal logic formalism containing constructors for modalities. In particular, the intuition is that measurements can be seen as tests and the evolutions of the system as programs.

How do we prove the correctness of an algorithm in this setting? Essentially, by proving that it is a validity of the logic. More precisely, if we are able to represent the systems we want to study as Modal Logic frames, we can express the correctness of an algorithm by proving that the formula encoding the algorithm is true at all states in all systems, i.e., is a validity of the corresponding class of Modal Logic frames. The key result that we need to apply the above line of reasoning is Soundness: we need to show that if a formula is provable in the logic (from some premises) then it is true in all states in all Modal Logic frames (satisfying the premises).

3.1 *LQP*

The logic LQP is an implementation of these ideas. Given a set of atomic propositions At and a set of atomic actions $AtAct$, the set of formulas $\mathcal{F}_{LQP}$ is built

by mutual recursion as follows:

$$\psi ::= p \mid \neg\psi \mid \psi \wedge \varphi \mid [\pi]\psi$$

where $p \in At$ and the action π is defined as

$$\pi ::= U \mid \pi^\dagger \mid \pi \cup \pi' \mid \pi;\pi' \mid \psi?$$

where $U \in AtAct$. The basic actions are meant to represent unitary transformations, while the tests $\psi?$ capture the measurement of a certain property. The composition of actions stands for the sequential composition of quantum gates or measurements, the dagger is the conjugate transpose (see previous section) and the nondeterministic union of actions is exploited to render the nondeterministic behaviour caused by the measurements.

The semantics of such a language is the following:

Definition 1. *Given a Hilbert space H, call L_H its lattice of linear subspaces. A* concrete quantum dynamic frame *is a tuple $\langle \Sigma_H, \{\xrightarrow{P_a?}\}_{a\in L_H}, \{\xrightarrow{U}\}_{U\in\mathcal{U}}\rangle$ such that*

(1) Σ_H is the set of all one-dimensional linear subspaces of H
(2) $\{\xrightarrow{P_a?}\}_{a\in L_H}$ is a family of quantum tests*, partial maps from Σ_H into Σ_H associated to the projectors of the Hilbert space H. Given $\overline{v} \in \Sigma_H$, the partial map $\xrightarrow{P_a?}$ is defined as $\xrightarrow{P_a?}(\overline{v}) = \overline{P_a(v)}$. The map is undefined if $P_a(v)$ is the zero vector.*
(3) $\{\xrightarrow{U}\}_{U\in\mathcal{U}}$ is a collection of partial maps from Σ_H into Σ_H associated to the unitary maps from H into H. As for projectors, the map $\xrightarrow{U}$ is defined as $\xrightarrow{U}(\overline{v}) = \overline{U(v)}$.

Call Γ_{CQDF} the class of all concrete quantum dynamic frames.

Definition 2. *An LQP-*model *M consists of a concrete quantum dynamic frame $\langle \Sigma_H, \{\xrightarrow{P_a?}\}_{a\in L_H}, \{\xrightarrow{U}\}_{U\in\mathcal{U}}\rangle$ and a valuation function $V : At \to \wp(\Sigma_H)$.*

Given an LQP-model, we define an interpretation of the actions and the satisfaction relation by mutual recursion.

Definition 3. *An* interpretation of the actions *in an LQP-model is a function $i : AtAct \to \{\xrightarrow{P_a?}\}_{a\in L_H} \cup \{\xrightarrow{U}\}_{U\in\mathcal{U}}$ such that*

- $i(U) \in \{\xrightarrow{U}\}_{U\in\mathcal{U}}$
- $i(\pi \cup \pi') = i(\pi) \cup i(\pi')$
- $i(\pi^\dagger) = i(\pi)^\dagger$
- $i(\pi;\pi') = i(\pi); i(\pi')$
- $i(\psi?) = \xrightarrow{P_a?}$ *where a is the span of the set $\{s \in \Sigma_H | M, s \vDash_{LQP} \psi\}$*

Where ; on the right-hand side is the composition of relations (partial functions in this case), $\cup$ is the union of relations and $\dagger$ is an analogue of the conjugate transpose for relations (see [7], Definition 11 p.27).

Definition 4. *Given a model M, a state s in the model and a formula $\psi \in \mathcal{F}_{LQP}$, the satisfaction relation $\vDash_{LQP}$ is defined as*

- $M, s \vDash_{LQP} p$ *iff* $s \in V(p)$
- $M, s \vDash_{LQP} \neg\psi$ *iff* $M, s \nvDash_{LQP} \psi$
- $M, s \vDash_{LQP} \psi \wedge \varphi$ *iff* $M, s \vDash_{LQP} \psi$ *and* $M, s \vDash_{LQP} \varphi$
- $M, s \vDash_{LQP} [\pi]\psi$ *iff for all* $(s, s') \in i(\pi)$ *we have* $M, s' \vDash_{LQP} \psi$

Theorem 2 (Theorem 3 in [5], p. 21). *There exist a proof system of LQP which is sound with respect to the class Γ_{CQDF}.*

We refer to [5] for the details of the proof system.

3.2 LQP^n

Nevertheless, the formalism of LQP is not enough to express quantum protocols. We need to express *locality*, that is, we need to express the fact that some quantum gates or measurements are performed locally, on certain subsystems. For this reason we develop an enhanced version of LQP, called LQP^n, able to capture locality in systems of n qubits.

Suppose given a natural number n. Set $N = \{1, \ldots, n\}$ and $I \subseteq N$. The syntax of LQP^n is the same as that of LQP plus the propositional constants

$$\top_I \,|\, 1 \,|\, +$$

and the constant actions $triv_I$. The new propositional constants are used to characterize local properties, while $triv_I$ is used to define local actions.

Definition 5. *Let H' be a Hilbert space of dimension 2 with basis $\{|1\rangle, |0\rangle\}$. Consider the Hilbert space $H := \otimes_{i=1}^{i=n} H'$ consisting of n copies of H' and call* n-partite quantum dynamic frame *the concrete quantum dynamic frame associated to H.*

Set $N = \{1, \ldots, n\}$. We write H_I to indicate the tensor product of the Hilbert spaces indexed by the indices in I. Thus in particular $H_N = H$.

Call Γ_{CQDF^n} the class of n-partite quantum dynamic frames.

Definition 6. *The satisfaction relation $\vDash_{LQP^n}$ contains that of LQP and is defined on the new formulas as*

- $M, s \vDash_{LQP^n} 1$ *iff* $s = \overline{\otimes_{i=1}^{i=n} |1\rangle_i}$
- $M, s \vDash_{LQP^n} +$ *iff* $s = \overline{\otimes_{i=1}^{i=n} |+\rangle_i}$
- $M, s \vDash_{LQP^n} \top_I$ *iff* $s \in \top_I^{\Sigma}$

Hence 1 and $+$ are used to denote specific states, while the last condition means that $\top_I$ is true at a state iff that state is I-separated.

Theorem 3 (Theorem 7 in [5], p.28). *There is a proof system of LQP^n which extends that of LQP and is sound with respect to the class Γ_{CQDF^n}.*

The correctness of a protocol can then be proved by encoding it in a formula of the logic and then showing that such formula is a validity.

Theorem 4 ([5] and [3]). *In the logic LQP^n we can give a formal correctness proof of the following algorithms: Teleportation, Quantum Secret Sharing, Superdense Coding, Entanglement Swapping and Logic Gate Teleportation.*

Unfortunately, with this logic we can only encode quantum protocols that succeed with probability 1. We will try to overcome this limitation in the second part of the next section.

4 Drawing the connection

We are naturally inclined to ask: is there a way to connect these two approaches?

Definition 7. *A* small category *is a category such that the collection of objects and the collection of maps are both sets. A* locally small category *is a category such that, for each pair of objects I, J, the collection of morphisms from I to J is a set.*

Definition 8. *Given a small category* $\mathbf{C}$ *and a functor* $U : \mathbf{C} \to \mathbf{Rel}$*, a* $(\mathbf{C}, U)$*-frame is a pair* $\langle W, Rel \rangle$ *such that:*

- $W := \bigcup \{U(I) | I \in \mathbf{C}_0\}$
- $Rel := \{U(f) | f \in \mathbf{C}_1\}$

Notice that if $\mathbf{C}$ *is small then* W *is the union of set-many sets, and thus is a set. Similarly, as there are set-many morphisms in* $\mathbf{C}$*, Rel will be a set.*

This procedure cannot be applied to locally small categories, if we want to have a set-sized carrier and set-many relations. However, we can give up the idea of obtaining a single Modal Logic frame from a category and have instead *a frame from every small subcategory*. Note that this entails having class-many frames. Hence from a big category we can still recover a class of frames, and study such class with modal logics.

We want to apply this construction to **FdHil** to recover the semantics of LQP and LQP^n. In this particular case, we are interested in having one frame for each physical system, because this was the original idea underlying LQP, thus we will consider the frames generated by the subcategories of **FdHil** with only *one* object. Notice that the procedure depends on the choice of a functor $\mathbf{FdHil} \to \mathbf{Rel}$; each functor produces a different class of frames, and therefore a different modal logic.

Consider the functor $S : \mathbf{FdHil} \to \mathbf{Rel}$ defined as:

$$\begin{aligned} H &\mapsto \Sigma_H \\ L : H \to V &\mapsto S(L) : \Sigma_H \to \Sigma_V \end{aligned}$$

where Σ_H is the set of one-dimensional linear subspaces of H and the functions $S(L)$ are the partial functions defined as $S(L)(\overline{v}) = \overline{L(v)}$. We now define a $(\mathbf{H}, S)$-frame, for $\mathbf{H}$ the full subcategory of **FdHil** containing only a single Hilbert space H.

Definition 9. *An $(\mathbf{H}, S)$-frame is a pair $\langle W, Rel\rangle$ such that*

- $W := \Sigma_H$
- $Rel := \{S(L) | L \in \mathbf{H}_1\}$

Hence in this case the carrier of the Modal Logic frame is the set of all one-dimensional subspaces of H. Alternatively, since such subspaces are in bijection with the unitary vectors that represent the states of a quantum systems, W is the set of all states of H.

Notice that the concrete quantum dynamic frame given by a Hilbert space H is a substructure of the corresponding $(\mathbf{H}, S)$-frame: the latter has all the partial functions corresponding to linear maps of type $H \to H$, the former only those corresponding to unitary maps and projectors. So if we interpret the programs in the syntax of LQP in the "right "way, that is, we send tests to the partial functions corresponding to projections and basic actions to the partial functions corresponding to unitary transformations, we get the same validities of the class Γ_{CQDF} in the language of LQP. This happens simply because all the additional relations that are in the $(\mathbf{H}, S)$-frame but not in the concrete quantum dynamic frame are not expressible in the language.

Theorem 5. *The logic of the class of $(\mathbf{H}, S)$-frames in the language $\mathcal{F}_{LQP}$ contains all the theorems of LQP. Similarly for LQP^n, when the class of frames is suitably restricted.*

4.1 Capturing probabilistic information

We can also adopt the same technique to obtain richer relational structures, for example structures containing probabilistic information. Consider $F : \mathbf{FdHil} \to \mathbf{Rel}$ defined as:

$$\begin{aligned} H &\mapsto A_H \\ L : H \to V &\mapsto F(L) : A_H \to A_V \end{aligned}$$

The set A_H is the set of functions $s_\rho : L_H \to [0, 1]$, where L_H is the lattice of closed linear subspaces of H, defined as

$$s_\rho(a) = tr(P_a \rho)$$

where P_a is the projector associated to the subspace a and ρ is a density operator on H.

A linear map $L : H \to V$ is sent to the partial function $F(L) : A_H \to A_V$ where

$$F(L)(s_\rho) = s_{\frac{L\rho L^\dagger}{tr(L\rho L^\dagger)}}$$

$F(L)(s_\rho)$ is not defined if $tr(L\rho L^\dagger) = 0$. Recall that the density operators are exactly the positive linear maps of trace 1. The operator $L\rho L^\dagger$ is still positive, and the denominator $tr(L\rho L^\dagger)$ ensures that it is an operator of trace 1. Therefore $\frac{L\rho L^\dagger}{tr(L\rho L^\dagger)}$ is again a density operator, so the function $F(L)$ is well defined.

Definition 10. *A $(\mathbf{H}, F)$-frame is a pair $\langle W, Rel\rangle$ defined as*

- $W := A_H$
- $Rel := \{F(L) | L : H \to H\}$

Hence in this case the carrier of the Modal Logic frame is the set of functions $s_\rho : L_H \to [0, 1]$ defined above. Such functions are associated to density operators on H, which represent *both pure and strictly mixed states* of the quantum system H. The set Rel is the collection of maps generated by all the linear maps of type $H \to H$.

The next question to address is: can we devise a logic to express the features of these richer frames? Consider the following syntax.

Given a set of atomic propositions At, build the set $\mathcal{F}$ by recursion[5]

$$\alpha ::= p \,|\, \sim \alpha \,|\, \alpha \curlywedge \beta$$

Then put $At' := \mathcal{F} \times [0, 1]$. We will indicate the pairs in At' as α^r. Now build the syntax as for LQP, by mutual recursion, but using At' as set of atomic propositions

$$\psi ::= \alpha^r \,|\, Pure \,|\, \neg\psi \,|\, \psi \wedge \varphi \,|\, [\pi]\psi$$

The set of programs Act is defined as for LQP:

$$\pi ::= U \,|\, \pi^\dagger \,|\, \pi \cup \pi' \,|\, \pi; \pi' \,|\, \psi?$$

Call this set of formulas $\mathcal{F}_{Prob}$.

The idea is to interpret the formulas of type α in subspaces, and to say that a mixed state s_ρ satisfies α^r if the probability of the subspace associated to α is exactly r according to s_ρ. The atomic constant $Pure$ is used to distinguish between pure and strictly mixed states. Given a $(\mathbf{H}, F)$-model we define a satisfaction relation for formulas in $\mathcal{F}_{Prob}$ as follows. First define a function $v : \mathcal{F} \to L_H$ by putting

- $v(p) = \sqcup\{a | s^a \in V(p)\}$
- $v(\sim \alpha) = v(\alpha)'$
- $v(\alpha \curlywedge \beta) = v(\alpha) \cap v(\beta)$

where $\sqcup$ is the supremum of the lattice L_H, $'$ is the complement and $\cap$ is the infimum. As for LQP, we define the function i associating a relation $R \subseteq A_H \times A_H$ to every program:

- $i(U) \in \{F(U) | U : H \to H\} \subseteq Rel$

[5] We use different symbols for the connectives not to get confused with the next step.

- $i(\pi^\dagger) = i(\pi)^\dagger$
- $i(\pi \cup \pi') = i(\pi) \cup i(\pi')$
- $i(\pi; \pi') = i(\pi) \circ i(\pi')$
- $i(\psi?) = F(P_{a'})$, where $a' = \sqcup\{a \in L_H | s^a \vDash_{Prob} \psi\}$

By mutual recursion the satisfaction relation is then:

- $M, s_\rho \vDash_{Prob} \alpha^r$ iff $s_\rho(v(\alpha)) = r$
- $M, s_\rho \vDash_{Prob} Pure$ iff $s_\rho \in Pure(A_H)$
- $M, s_\rho \vDash_{Prob} \neg\psi$ iff $M, s_\rho \nvDash_{Prob} \psi$
- $M, s_\rho \vDash_{Prob} \psi \wedge \varphi$ iff $M, s_\rho \vDash_{Prob} \psi$ and $M, s_\rho \vDash_{Prob} \varphi$
- $M, s_\rho \vDash_{Prob} [\pi]\psi$ iff for all $(s_\rho, s_{\rho'}) \in i(\pi)$ we have $M, s_{\rho'} \vDash_{Prob} \psi$

This language can be enhanced further in order to express locality, mimicking the move from LQP to LQP^n. We refer to [7] p.73 for details.

The first result that we want to obtain is the preservation of the correctness proofs casted in the proof system LQP^n. This can be shown in two steps. First, the language $Prob$ is an extension of the language of LQP (similarly for LQP^n), hence we have a translation of the syntax of LQP into $Prob$. Essentially, the atomic propositions of LQP (LQP^n) are encoded in atomic propositions with superscript 1, while for the other cases the atomic proposition $Pure$ is used to ensure that we encode $LQP(LQP^n)$-formulas in the right way. Second, Modal Logic frames given by F can be turned into the corresponding Modal Logic frames given by S.

Proposition 1 ([7] Proposition 14, p.69). *There is a natural transformation $\delta : S \to F$ defined componentwise as*

$$\delta_H = \{(a, s^a) | a \in \Sigma_H, s^a \in A_H\}$$

recall that, being a morphism in **Rel**, *δ_H is a relation of type $\Sigma_H \to A_H$. The relation δ_H associates every one-dimensional linear subspace to the (function associated to) corresponding density operator.*

This natural transformation gives us a way to transform models obtained from the functor F into models obtained from S, basically forgetting the additional information preserved by F. Piecing together there two facts we can show, using the soundness of LQP (respectively, LQP^n):

Theorem 6 ([7] Theorem 15 and 16, pp.71-75). *Upon translation, all the theorems of LQP (LQP^n) are validities of the class of $(\mathbf{H}, F)$-frames (given by compound systems) in the language of $Prob$.*

5 Conclusions

We have seen how from the category **FdHil** and the functor S we can obtain the class of Modal Logic frames for LQP and LQP^n. This constitutes the link between the two research programs that we have considered.

Our second case study, the functor F, has highlighted the possibility to obtain a richer semantics. We have proposed a logic for such class of Modal Logic frames, and proven that it is an improvement with respect to LQP and LQP^n, in the sense that it has more expressive power and it contains all the (translations of) the theorems of LQP and LQP^n.

Acknowledgments

The author would like to thank the anonymous referees of ESSLLI Student Session for their useful comments and suggestions. This paper is extracted from the Master Thesis of the author (see [7]), which was written under the supervision of Alexandru Baltag. My gratitude goes to him and to the members of the Thesis Committee, who helped me in improving the material contained in this article.

References

1. Abramsky, S., Coecke., B.: A categorical semantics for quantum protocols. In: In: Proceedings of the 19th Annual IEEE Symposium on Logic in Computer Science (LiCS'04), IEEE Computer Science. Press. Arxiv:quant-ph/0402130 (2004)
2. Abramsky, S., Coecke., B.: Categorical quantum mechanics. Handbook of Quantum Logic Vol. II (2008)
3. Akatov, D.: The Logic of Quantum Program Verification. Master thesis, Oxford University Computing Laboratory (2005)
4. Baltag, A., Smets., S.: Complete Axiomatizations for Quantum Actions. International Journal of Theoretical Physics 44(12), 2267–2282 (2005)
5. Baltag, A., Smets., S.: LQP, The Dynamic Logic of Quantum Information. Mathematical Structures in Computer Science 16(3), 491–525 (2005)
6. Blackburn, P., de Rijke, M., Venema, Y.: Modal Logic. Cambridge Tracts in Theoretical Computer Science, Cambridge University Press (2002)
7. Cinà, G.: On the connection between the categorical and the modal logic approaches to Quantum Mechanics. Master thesis, ILLC (2013)
8. Lane, S.M.: Categories for the Working Mathematician. 2nd edition. Springer (1998)
9. Nielsen, M., Chuang, I.: Quantum Computation and Quantum Information. Cambridge University Press (2010)

Relevance, Relevance and Relevance in Epistemology and Logic *

Peter Hawke

Philosophy Department, Stanford University
Stanford, CA 94305, USA
phawke@stanford.edu

Abstract Our topic is the relevant alternatives (RA) approach to knowledge attribution. In terms of epistemic logic, we here understand such a theory as claiming that knowledge-that is not closed under entailment. We offer three competing logical semantics for RA theory, inspired by important informal discussions in the literature due to Dretske, Schaffer and Yablo. We comment on the distinctive logics of relevance that emerge and argue that the third of our candidates best meets certain important desiderata.

1 Introduction

The relevant alternatives theory of knowledge attributions[1] can be summarized as follows:

> When P is true, "S knows that P" is true just in case S has *ruled out* all of the *relevant alternatives* to P.

RA theory is an *abstract* theory. A more particular theory may be generated under this umbrella by specifying exactly what is meant by the following key terms: *ruled out*, *relevant* and *alternative*. Of these, 'relevance' is the most crucial and the most vexed.

RA theory has many interesting features, a number of which play a role in this paper. From a philosophical point of view, RA theory finds motivation in its distinctive approach to skeptical and under-determination problems (more on this later). From the perspective of logic, RA theory presents an intriguing point of contact between epistemic logic and epistemology: RA theory lends itself to precise formal characterization, and raises interesting questions concerning the

* I wish to thank Johan van Benthem, Michael Bratman, David Hills, Wes Holliday, Krista Lawlor and three anonymous referees for helpful comments on the material in this paper.

[1] RA theory is generally presented in the philosophy literature as a theory of *knowledge*, not as a meta-linguistic theory of knowledge attributions. We will not detain ourselves considering the difference, if any, that this change in perspective brings about.

logic of knowledge claims. In particular, the issue as to whether knowledge is *closed under (known) entailment* is pertinent (more on this later).

In this paper, we work with a certain conception of RA theory: as the commitment that the *knowledge-that* operator is not *closed under entailment*. We aim to contribute to the project of spelling out an adequate account of such an RA theory[2].

To this end, we have two particular goals for the paper, to be achieved simultaneously. First, we aim to simply survey and clearly distinguish three different approaches to capturing our conception of RA theory: what we call the *worlds-first approach*, the *question-first approach* and the *topic-first approach*. Altogether, this accounts for major strands in the (informal) epistemology literature associated with Dretske [5,6], Lewis [11], Schaffer [15] and Yablo [16,18] (though many precise details are, for better or worse, our own). We emphasize that these strands can fruitfully be captured in a logical framework, giving rise to logics of knowledge and subsidiary epistemic notions. Notably, each framework represents a peculiar conception of relevance, with associated logical properties. Space precludes a detailed logical study, but we remark on the distinctive *logic of relevance* that emerges in each framework. While we borrow many ideas from the existing epistemology and epistemic logic literature, we hope that our exact syntheses of these ideas (and peculiar focus on e.g. relevance as an object of logical study) is new.

Our second particular goal is more philosophical, though with a logical dimension: we wish to tentatively provide reasons to *favor* one framework. We introduce desiderata for an adequate framework to fulfil and briefly argue that the most promising on offer is what we call the *topic-first approach*, which is exemplified here by a logic that is a hopefully novel amalgam of ideas due to Yablo [16,18], Lewis [11] and the awareness literature in epistemic logic [7][2, Ch.5]. We suggest this framework has attractive tools for allaying perennial concerns connected to epistemic skepticism and epistemic closure failure.

2 Preliminaries

The logical frameworks we introduce will each be a variation of the following basic framework, based on standard ideas in the literature.

[2] For those readers familiar with the literature, it should be emphasized that we recognize that not every theory in the epistemology literature that travels under the moniker 'relevant alternatives theory' involves the claim that knowledge-that is not closed under entailment. For an influential 'RA theory' that preserves closure, see for instance [12]. According to a liberal account of RA theory, our present concern can be understood as focusing on a certain important *sub-class* of RA theories. That we here *identify* RA theory with the rejection of closure is essentially an expository device. We explain our rationale in section 3.1.

2.1 An epistemic language

We work throughout with a fixed set of atomic proposition symbols and the language $\mathcal{L}_E$:

$$\phi ::= \quad p \mid s \mid \phi \vee \phi \mid \neg\phi \mid K\phi \mid I\phi \mid R\phi \mid \Box\phi$$

s is a special proposition letter acting as a skeptical hypothesis e.g. "I am a brain-in-a-vat". The remaining boolean connectives can be defined as usual. We write $\phi \dashrightarrow \psi$ to indicate $\Box(\phi \rightarrow \psi)$. $K\phi$ is intended to mean "S has the knowledge that ϕ". $I\phi$ is intended to mean "S has the semantic information that ϕ". $R\phi$ is intended to mean "ϕ is relevant to S's epistemic state". $\Box\phi$ is intended to mean "ϕ is necessary".

We sometimes work with a variation of this language where the R operator is two-place: we write $R(\phi, \psi)$ to mean "ψ is relevant to S's epistemic state with respect to ϕ".

2.2 Basic epistemic models

Definition 1. *A* basic epistemic model $\mathcal{M}$ *is a tuple* $\langle W, @, \sim, B, \mathbf{s}, V\rangle$ *where:* W *is a finite set of* possible worlds *(finitude is for technical convenience);* $@ \in W$ *is a distinguished* actual world*;* $\sim$ *is an equivalence relation on* W *called the* indistinguishability relation*;* $B \subseteq W$ *is a set of skeptical (brain-in-vat) worlds such that* $s \in S$ *implies that* $s \sim @$*;* V *assigns proposition letters to the members of* W*; and* $\mathbf{s} \in B$ *is a distinguished* skeptical world *such that* $V(\mathbf{s})$ *is the complement of* $V(@)$*.*

Our modeling aspirations: the indistinguishability relation models the *semantic information* at our agent's disposal: if $w_1 \sim w_2$ then the agent's information cannot distinguish w_1 and w_2. The information we have in mind is that received by the agent in interacting with the world. We say that a world w_2 is *eliminated* with respect to w_1 if $w_1 \nsim w_2$.

We flag an idealization that we make throughout (we will flag a second idealization in section 3.3): our agent has perfect *uptake* of the information available to her. Our agent considers a possible world as a live possibility if and only if that world is compatible with the information transmitted to her by the world. Our agent is not subject to misinformation, imperfect uptake etc. Worlds indistinguishable from actuality may therefore be regarded as our ideal agent's *belief set* or *evidence*.

B is a set of *skeptical scenarios*. For ease, one may think of these worlds as the epistemologist's *brain-in-vat* worlds[3]: here, the agent is a bodiless brain-in-vat whose every experience is a fabrication by alien scientists. We presume the

[3] One could replace this radical example with another example of a skeptical possibility found in the literature: the possibility that the animal the agent *in fact* identifies correctly is a zebra is a cleverly disguised mule; the possibility that the object the agent *in fact* correctly identifies as a goldfinch is in fact a fake mechanical toy etc. It is convenient for technical reasons to work with radical skepticism, however.

skeptical worlds *cannot* be eliminated by the agent's (actual) information. In world **s**, the agent is massively deceived: thanks to the inclusion of such a world, we have $\mathcal{M}, @ \not\models Ip$ for every proposition letter p[4].

2.3 Basic semantics

We describe the satisfaction relation $\models$ for a (here) uncontentious fragment of our language.

Definition 2. *Given epistemic model* $\mathcal{M}$*:*

- $\mathcal{M}, w \models p$ *iff* $p \in V(w)$.
- $\mathcal{M}, w \models s$ *iff* $w \in B$.
- $\mathcal{M}, w \models \neg\phi$ *iff* $\mathcal{M}, w \not\models \phi$.
- $\mathcal{M}, w \models \phi \vee \psi$ *iff* $\mathcal{M}, w \vDash \phi$ *or* $\mathcal{M}, w \vDash \psi$.
- $\mathcal{M}, w \models I\phi$ *iff* $\mathcal{M}, v \models \phi$ *for every* v *such that* $w \sim v$.
- $\mathcal{M}, w \models \Box\phi$ *iff* $\mathcal{M}, v \models \phi$ *for every* $v \in W$.

3 Relevant Alternatives Theory

3.1 Nature of RA theory

We focus our discussion of RA theory with two natural stipulations: an *alternative* to a ϕ is any ψ that entails $\neg\phi$. Formally: $\psi \dashv \neg\phi$. Second, ϕ is *ruled out* for agent S just in case S knows that $\neg\phi$. Formally: $K\neg\phi$.

This is not the only way to try to understand what "ruling out" a proposition comes to, and other approaches are taken in the literature. We consider our current approach to be particularly natural. Is it not strange to say that one has *ruled out* that Alan Turing was born in the 19th century, yet hesitate to say that one *knows* that he was *not* born in the 19th century? Is it not strange to say that one *knows* that Alan Turing was born in the 20th century, yet hesitate to say that it is *ruled out* that he was *not* born in the 20th century? The *value* of a true knowledge claim is precisely that it *rules out* alternatives.

On our conception, then, an RA theory is one such that: it is possible that both P is known and A is an alternative to P, yet $\neg A$ is not known. An RA theorist claims that *closure under entailment* does not hold:

$$\text{Closure under entailment:} \quad \vDash (\phi \dashv \psi) \rightarrow (K(\phi) \rightarrow K(\psi)).$$

Such a failure occurs, according to RA theory, exactly when alternative $\neg\psi$ is *irrelevant*. Thus, we expect a satisfactory RA account to entail the following validity:

$$\vDash (K(\phi) \wedge (\psi \dashv \neg\phi) \wedge \neg K(\neg\psi)) \rightarrow \neg R(\psi)$$

One important goal of an RA theorist, on our conception, is to give a satisfying semantics for both the K operator and the R operator that ensures this validity.

[4] This does not mean our agent has no semantic information. For instance, the agent might have conditional information $\mathcal{M}, @ \vDash I(\neg s \rightarrow h)$.

3.2 Motivation

Why be an RA theorist? An important motivation is the allure of a certain kind of strategy against philosophical skepticism. "I know that I have hands". "I don't know that I am not a handless brain-in-a-vat". The former seems unobjectionably true. Yet the latter, worryingly, also strikes many as true (hence the force of skeptical arguments). As observed by Dretske [5], it is tempting to accept both. But why then does our lack of knowledge about the latter not affect our knowledge of the former? The RA theorist's answer: claims about brains-in-vats are *irrelevant* to assessing everyday knowledge claims. Skeptical hypotheses are *far-fetched*; or change the *issue* or *subject matter* from the banal to the sublime; or introduce unusually strict epistemic standards. Or so forth.

3.3 Desiderata

What does a satisfactory RA theory look like? We propose three general desiderata:

(1) **Skeptical hypotheses:** Consider "I am a handless brain-in-vat" ($\neg h \wedge s$) (where h holds at actuality). If h is known, then it should not follow that $\neg(\neg h \wedge s)$ is known (as per our motivation above). Formally:

$$\not\models Kh \rightarrow K(h \vee \neg s)$$

(2) **Closure:** while RA theory, on our conception, is committed to a general kind of closure failure, it seems important to avoid specific instances of closure failure. Indeed, the first of the principles below is mandatory for the RA theorist, and the following two principles find significant endorsement in the literature[5]:

   (a) Closure under entailment restricted by relevance. Formally:

$$\models (K(\phi) \wedge (\psi \mathrel{-\!\!\!\triangleright} \neg\phi) \wedge R(\psi)) \rightarrow K(\neg\psi)$$

   Note that this is just an alternative statement of the desideratum mentioned in section 3.1.

   (b) Closure under *known* entailment. Formally:

$$\models K(\phi \mathrel{-\!\!\!\triangleright} \psi) \rightarrow (K\phi \rightarrow K\psi)$$

[5] For supporters of the preservation of closure under known entailment, see [12] and [10], among many others. It should be noted, however, that it is rare to hear support for the preservation of closure under known entailment, while simultaneously giving up closure under entailment (in particular, the authors just cited are not RA theorists in the sense of the present paper). Support for these breeds of closure generally comes as a package. We challenge this dogma momentarily. We may also note that [10] and [16,18] explicitly point out the intuitiveness of conjunction elimination.

(c) Conjunction elimination. Formally:

$$\vDash K(\phi \wedge \psi) \rightarrow K(\psi)$$

So, an RA theorist will want to reject $K(h \wedge \neg(\neg h \wedge s))$ if she wishes $\neg K(\neg(\neg h \wedge s))$ to hold i.e. she must reject ordinary knowledge of being handed and not a handless brain-in-vat if she wishes it to hold that we do not ordinarily know that we are not handless brains-in-vats.

(3) **Properties of relevance:** "relevance" is a somewhat technical notion for the RA theorist. Nevertheless, the appropriateness of this terminology trades on certain intuitions. For instance: it is strange to say that $\neg P$ is irrelevant to our assessment of S, yet P *is* relevant[6]. It is more natural to say: *whether* P *is the case* is (ir)relevant to our assessment of S. Formally: $\vDash R\phi \leftrightarrow R(\neg\phi)$[7].

Two immediate objections to the second listed desideratum might have struck the reader. First, it might strike one that closure under known entailment is obviously *invalid* for ordinary agents: an ordinary person might know P and know that P implies Q yet fail to put "two-and-two together". Second, it might strike one that it is unnatural to simultaneously fix as desiderata *both* the rejection of closure under entailment *and* the acceptance of closure under *known* entailment. Indeed, the tendency in the philosophy literature is to either endorse both or reject both.

We respond to the first criticism by emphasizing a second idealization concerning the agent we model with our formal framework: as it typical in the epistemic logic literature, we deliberately seek to model an agent that is an *ideal reasoner with no computational bounds*. If knowledge of P is 'available' to such an agent by way of "putting two-and-two together" using her existing knowledge, then this agent knows P[8]. Why care about this kind of agent? First, note that such an agent is perfectly susceptible to skeptical worries concerning empirical knowledge, so the move to idealization does not lose our focus on the problem at hand. Second, this idealization allows us to imbue the above formal principle of closure under known entailment with intuitive significance: as originally pointed out by [10], if this principle fails for ideal reasoners, then that means that an

[6] This consequence is connected to the above desideratum concerning skeptical hypotheses. As RA theorists, we are committed to the idea that the possibility that one is a handless brain-in-vat is irrelevant when considering a true ordinary knowledge claim that one has a hand. It seems odd to add that it *is* relevant that one is *not* a handless brain-in-vat. For, for one thing, relevant propositions are 'serious candidates' for knowledge from the RA perspective, and discomfort with the thought that one could possibly *know* that one is not a handless brain-in-vat is exactly the sort of consideration that drove us to RA theory.

[7] Further logical constraints on R might seem attractive e.g. $\vDash K\phi \rightarrow (R\phi \wedge R(\neg\phi))$.

[8] Note then that we are not concerned to address the problem of logical omniscience in our logical frameworks. To accept the failure of closure under entailment, on the basis of the response to skepticism from section 3.2, is to accept this failure even for *ideal reasoners*. This is worth keeping in mind when we deploy some tools from the awareness logic literature, first developed to deal with logical omniscience.

ordinary agent cannot always conclude that a consequence of her knowledge - *known* by her to in fact be a deductive consequence - is itself knowledge. This is a highly counter-intuitive claim, flying in the face of the seemingly obvious fact that deduction is an infallible tool for the extension of knowledge.

In response to the second criticism: we believe it is a mistake to think that the reasons for supporting both kinds of closure are perfectly symmetric. First, we are interested in exploring RA theory understood as the rejection of closure under entailment (as one natural reading of the RA slogan). On the face of it, this reading of the RA slogan holds no immediate implication for closure under *known* entailment, so our starting point does not *by itself* weigh against closure under known entailment. Second, there are intuitive considerations that support closure under known entailment that do not seem to directly bear on mere closure under entailment: namely, the connection to *acquiring knowledge under deductive reasoning* mentioned in the previous paragraph.

A somewhat persuasive argument for rejecting both forms of closure, however, is that counter-examples to one seem to be easily transformed into counter-examples for the other. In section 3.2, we mentioned that one way to escape the skeptical argument is to diagnose it as pointing to a counter-example to the claim that knowledge is closed under entailment. However, since we apparently *also* know that being a brain-in-vat entails not having hands, it might seem that we likewise have a counter-example to closure under *known* entailment. The only way to deny this last conclusion is to deny that (at least in 'ordinary contexts') we know that being a brain-in-vat entails not having hands. Since this entailment represents a simple, analytic truth, this commitment is hard to swallow. Nevertheless, it can be checked that our preferred logic in the next section makes *exactly* this commitment: in fact, formulae of the form $\neg K(\phi)$, where ϕ is a *tautology*, are satisfiable in that logic. This might seem a particularly strange outcome, given that we focus attention on agents that are *ideal reasoners*. The explanation, however, strikes us as simple and natural in the context of the rationale for that logic: there is no reason to suppose that every tautological sentence is *relevant* in an epistemic context when we adopt the topic-first approach, for that sentence might mention irrelevant subject matter (we expand on this point shortly). Then, if the irrelevance of a sentence intuitively amounts to it being 'properly ignored' in that epistemic context (Lewis' terminology), then it is not particularly unnatural to say that it is *not known*. The diagnosis of this lack of knowledge, of course, is *not* a lack of evidence: it is simply that ϕ is not a suitable *candidate* for knowledge in that context[9].

[9] And, as the reader might investigate for herself, our logical framework still assigns a special role to tautological and other logical truths for our ideal agent: namely, these statements are *necessarily* known by our agent *if* they are relevant. Of course, some readers will find this treatment of the knowledge of logical truths to be overly elaborate and artificial. Such readers will inevitably consider this feature of our preferred logic as a cost to adopting that approach, to be weighed against its benefits.

4 Accounts of Relevance and Knowledge

We now briefly discuss three proposals for an adequate RA theory, and consider the unique account of relevance attached to each. We point out difficulties for all but the last.

4.1 Worlds-first approach

Consider the following leading idea: primitively, it is *possible worlds* that are classifiable as relevant or irrelevant. What makes a possible world relevant? One proposal of many: for a possible world to be relevant, it needs to be *sufficiently similar* to the actual world.

Once we have the *relevance of worlds* on the table, we can define *relevance of a proposition* as follows: ϕ is relevant just in case it holds at some relevant world.

How to develop this idea into an RA theory? We build on the lead of Dretske [5][6], Nozick [13] and an ensuing systematic study by Holliday [9]. We equip our epistemic models with a similarity ordering on worlds. We can then say that the agent knows that ϕ just in case the worlds *most similar* to actuality in which $\neg\phi$ holds are eliminated by the agent's information. A corresponding notion of relevance: relative to ϕ, a world is relevant if it is as or more similar to actuality as the nearest $\neg\phi$ worlds.

Definition 3. *A WF-model $\mathcal{M}_{\preceq}$ is a basic epistemic model supplemented with a set of total pre-orderings on W - one $\preceq_w$ for each world - such that $w \preceq_w v$ for every $v \in W$.*

We may now expand our semantics.

- $\mathcal{M}_{\preceq}, w \vDash R(\phi, \psi)$ iff there exists a world u such that $\mathcal{M}_{\preceq}, u \vDash \psi$ and $u \preceq_w min_\phi$ where min_ϕ is any world such that $\mathcal{M}_{\preceq}, min_\phi \vDash \neg\phi$ and there is no world v such that $v \prec min_\phi$ and $\mathcal{M}_{\preceq}, v \vDash \neg\phi$.
- $\mathcal{M}_{\preceq}, w \vDash K\phi$ iff ($\mathcal{M}_{\preceq}, w \vDash \phi$ and) $w \nsim min_\phi$ for every world min_ϕ where: $\mathcal{M}_{\preceq}, min_\phi \vDash \neg\phi$ and there is no world u such that $u \prec min_\phi$ and $\mathcal{M}_{\preceq}, u \vDash \neg\phi$.

It may be checked that this generates an RA theory. To see that closure under entailment fails, construct a model where the nearest $\neg h$ world (to actuality, where $h \wedge \neg s$ holds) is eliminated, but the nearest $\neg h \wedge s$ world is not so eliminated.

The logic of relevance generated by this approach is of interest. Distinctive validities: $\vDash (\psi_1 \multimap \psi_2) \rightarrow (R(\phi, \psi_1) \rightarrow R(\phi, \psi_2))$ and $\vDash (\phi_1 \multimap \phi_2) \rightarrow (R(\phi_2, \psi) \rightarrow R(\phi_1, \psi))$.

Nevertheless, there are drawbacks to the current proposal.

Proposition 1. *According to the current semantics, closure under known entailment is not valid and conjunction elimination is not valid.*

Proposition 2. *According to the current semantics, $\nvDash R(\phi, p) \leftrightarrow R(\phi, \neg p)$.*

(Proving the propositions in this paper is a straightforward exercise we leave to the reader.)

4.2 Question-first approach

Another leading idea, forcefully pursued by Schaffer [15]: knowing that P is to implicitly know P *as the answer* to some or other (background) question. Whether one knows P or not may therefore depend on what question is being addressed. One may know that one has a hand against the backdrop of the question "is my hand still attached to my wrist?" but not against the backdrop of "am I a handless brain-in-vat?". A natural account of relevance arises: a proposition P is *relevant* relative to question Q just in case P is a (more or less specific) *answer* to Q. One way a proposition may be irrelevant: that proposition amounts to a denial of a *presupposition* behind the current question (that one is not a brain-in-vat is plausibly a presupposition behind "is my hand still attached to my wrist?").

Formal approaches to dealing with questions have been developed in a tradition emanating from Hamblin [8] [1] [2, Ch.6] [4]. Along these lines, we understand a question Q as a *set of mutually disjoint propositions i.e. most general answers*. An answer to Q is a subset of a member of Q; a partial answer is a *union* of such subsets. We do not suppose that these propositions are exhaustive: a proposition that entails $\bigcup Q$ is called a *presupposition of* Q.

Definition 4. *A QF-model $\mathcal{M}_Q$ is a basic epistemic model supplemented with a question Q: a set of mutually disjoint subsets of W.*

Our extended semantics:

- $\mathcal{M}_Q, w \vDash R\phi$ iff $[[\phi]] = A$ where $A \subseteq \bigcup Q$ and $[[\phi]]$ is the set of worlds in $\mathcal{M}_Q$ where ϕ holds.
- $\mathcal{M}_Q, w \vDash K\phi$ iff $\mathcal{M}_Q, w \vDash R\phi$ and $\mathcal{M}_Q, u \vDash \phi$ holds for every world u such that $w \sim u$ and $u \in \bigcup Q$ (and $\mathcal{M}_Q, w \vDash \phi$).

Note that we deploy an idea inspired by Fagin and Halpern's logic of awareness here [7]: $K\phi$ holds only if ϕ is relevant (for Fagin and Halpern: explicit knowledge entails *awareness*, as opposed to relevance). The motivation: irrelevance of ϕ, intuitively, means that ϕ is 'properly ignored' in the current epistemic context (i.e. against the backdrop of the *question* in play) and so is not a proper object for knowledge[10].

Thus, to know P is for P to be the least specific true partial answer (to the current question) that is uneliminated by the available information. Again, it may be easily checked that we have an RA theory on our hands: to see that closure under entailment fails, simply consider a model where $\neg s$ is a presupposition to the question in question. Thus: $h \vee \neg s$ (for h known at @) is a non-answer to Q, and so cannot be known.

Our current logic of relevance is distinguished by the following validity: $\vDash (\phi \multimap \psi) \rightarrow (R(\psi) \rightarrow R(\phi))$.

[10] Though this stipulation seems to us natural, note that this goes beyond Schaffer's explicit account.

Proposition 3. *Conjunction elimination is not valid according to the current semantics.*

Proposition 4. *On the current semantics,* $\not\models Rp \leftrightarrow R(\neg p)$.

4.3 Topic-first approach

A final leading idea, building on recent pursuits by Yablo [16][11]: to know P is to know it against the backdrop of a fixed *subject-matter* or *topic of inquiry/discussion*. On this view, roughly, the reason that one can know that one has hands (in ordinary circumstances), yet not know that one is not a handless brain-in-vat, is that discussion of an everyday concern involves a subject-matter that does not incorporate deep philosophical distinctions.

The distinction between a background *question* and a background *subject-matter* might seem blurry. Here is a useful contrast: an answer to a question can be *more specific than necessary*: one can successfully answer "which road leads to Rome?" with extremely detailed directions. On the other hand, to fix a subject-matter is to *preclude* discussion of more *specific* distinctions: if the topic of inquiry is how many *people* live in Rome, we purposefully neglect the finer grained topic of how many people *and rats* live in Rome. To 'bring up' the number of rats is to introduce *new* subject-matter into the conversation.

We model a topic syntactically as a *finite set of atomic sentences* T (for "topic")[12]. Our immediate inspiration for this syntactic approach is the awareness logic of Fagin and Halpern [7] (though also see the *relatedness logic* surveyed in *cf.* [3, Ch.5]). This induces an equivalence relation $\approx_T$ on our space of worlds (bringing us in line with Lewis' account of subject matter in [11]): $u \approx_T v$ just in case u and v agree on the valuation of every atom in T. This equivalence relation in turn induces a *partition* on W: call the members of this partition "small worlds"[13]. To know relevant P, on our current conception, is to have "eliminated" all of the small worlds in which P does not hold. As Yablo observes, "elimination" of a small world S should not be understood as each member of S being incompatible with the agent's information. For then small worlds which contain skeptical scenarios *cannot* be "eliminated". Rather, we should model elimination of S in terms of a select *subset* of members of S being incompatible with the agent's information. We cash out this idea by borrowing from Dretske[14]: to

[11] One source of Yablo's inspiration is Lewis [11]. Another philosopher who has taken up similar considerations is Yalcin [17].

[12] This is a departure from Yablo's own account. One technically significant difference is that a subject matter for Yablo does not necessarily correspond to an *equivalence relation* between worlds, but rather a *similarity relation* i.e. a reflexive and symmetric relation. Yablo's account of subject matter is highly nuanced but aslo leads his theory into some difficulties, we believe, though we do not elaborate here. We comment only that we view both features as reasons to present the variation of his approach outlined in this section.

[13] We adopt this terminology from Savage [14, Ch.5].

[14] Yablo, for his part, turns to Nozick for a solution.

"eliminate" S is to have information that is incompatible with the *nearest* worlds to actuality contained in S (following Dretske, we might say: the information is a *conclusive reason* to reject S).

We define the *relevant sentences* R_T as follows: $\phi \in R_T$ just in case every proposition letter occurring in ϕ is a member of T.

Definition 5. *A TF-model $\mathcal{M}_T$ is a basic epistemic model supplemented with a subject-matter T and a set of world-relative total pre-orderings, denoted $\preceq_w$, such that $w \preceq_w u$ for all u. A cell C in the partition induced by T is eliminated at w just in case: for every u in the set of nearest worlds to w in C, according to $\preceq_w$, we have that $w \nsim u$.*

Our extended semantics:

- $\mathcal{M}_T, w \vDash R\phi$ iff $\phi \in R_T$.
- $\mathcal{M}_T, w \vDash K\phi$ iff $\mathcal{M}_T, w \vDash R\phi$ and every cell in the partition induced by T that contains a $\neg\phi$ world is eliminated (and $\mathcal{M}_T, w \vDash \phi$).

In short: ϕ is known just in case ϕ is true, 'on-topic' and the agent's information is a conclusive reason to reject every possible way things could be *with respect to relevant subject matter T* in which ϕ is false. Again, we can check that closure under entailment fails: simply consider a model where s is not part of the subject matter, and so $h \vee \neg s$ cannot be known at the actual world @, even when h is known (for the cell containing skeptical world **s** to be conclusively rejected by the agent's information, set it so that no nearest $\neg h$-world to actuality is a s-world).

Some characteristic validities for the current account of relevance:
$\vDash R(\phi \wedge \psi) \rightarrow R(\phi)$ and $\vDash R(\phi) \leftrightarrow R(\neg\phi)$.

Proposition 5. *Our current semantics satisfies our immediate desiderata: the generated RA theory deals adequately with skeptical hypotheses, closure and the logic of relevance.*

5 Conclusion

We have made further progress in the ongoing project of bringing epistemology and epistemic logic closer together, by surveying and evaluating a series of proposed logical semantics for knowledge claims, inspired by major strands in the informal philosophical literature. In particular, we have offered a semantics - a topic-first approach, inspired by Yablo - that meets our immediate desiderata. Further investigation is required.

References

1. Belnap, N., T.B. Steel, *The Logic of Questions and Answers*, Yale University Press (1977)
2. van Benthem, J., *Logical Dynamics of Information and Interaction*, Cambridge University Press (2011).
3. Burgess, J.P., *Philosophical Logic*, Princeton University Press (2009)
4. Ciardelli, I., Groenendijk, J., Roelofsen, F. "On the Semantics and Logic of Declaratives and Interrogatives", *Synthese*, pp. 1-40 (2013)
5. Dretske, F., "Epistemic Operators", in *The Journal of Philosophy* 67(24), pp.1007-1023 (1970)
6. Dretske, F., "The Pragmatic Dimension of Knowledge", in *Philosophical Studies* 40, pp.363-378, Kluwer Academic Publishers (1981)
7. Fagin, R., J.Y. Halpern, Y., "Belief, Awareness and Limited Reasoning", in *Artificial Intelligence 34*, pp. 39-76 (1988)
8. Hamblin, C. L., "Questions in Montague English", *Foundations of Language* 10 (1), pp. 41-53 (1973)
9. Holliday, W., "Knowing What Follows", PhD Dissertation, Stanford University (2012)
10. Kripke, S., "Nozick on Knowledge", in *Philosophical Troubles: Collected Papers, Volume 1*, Oxford University Press (2011)
11. Lewis, D., "Relevant Implication", in *Theoria* 54, pp.161-174 (1988)
12. Lewis, D., "Elusive Knowledge", in *Australasian Journal of Philosophy* 74, 4, pp.549-567 (1996)
13. Nozick, R., *Philosophical Explanations*, Part 3, pp.167-247 (1981)
14. Savage, L.J., *The Foundations of Statistics*, Dover Publications (1972)
15. Lewis, D., "From Contextualism to Contrastivism", in *Philosophical Studies* 119, pp.73-103 (2004)
16. Yablo, S., 2012 Locke Lectures at Oxford University (2012)
17. Yalcin, S., "Nonfactualism About Epistemic Modality", in *Epistemic Modality*, A. Egan, B. Weatherson eds., Oxford University Press (2011)
18. Yablo, S., *Aboutness*, Princeton University Press (2014)

A Bimodal Provability Logic

Paula Henk

Institute for Logic, Language and Computation (ILLC),
University of Amsterdam
P.Henk@uva.nl

Abstract We introduce a nonstandard provability predicate $\triangle$ whose dual $\triangledown$ (with $\triangledown A := \neg\triangle\neg A$) is a unary supremum operator in the lattice of interpretability degrees of finite extensions of PA. As it turns out, the principles of the provability logic GL are valid for $\triangle$. We introduce a joint provability logic GLT for $\triangle$ and $\Box$ (the standard provability predicate of PA), as well as a suitable class of Kripke frames.

1 Introduction

The concept of *relative interpretability* was first introduced and carefully studied by Tarski, Mostowski and Robinson ([1]). Roughly speaking, a theory T interprets a theory S, we write $T \rhd S$, if there is some translation (preserving logical structure) from the language of S to the language of T, such that the translation of every theorem of S is a theorem of T. The notion of interpretability can be used to make precise the idea that one theory is at least as strong as another, even though their languages might be different[1]. The relation of interpretability is a preorder on theories; the equivalence classes of the induced equivalence relation (of mutual interpretability) are called *degrees*. Consider the collection V_{PA} of degrees of finite extensions of Peano Arithmetic (PA). The structure[2] $(\mathsf{V}_{\mathsf{PA}}, \rhd)$ is a distributive lattice ([2]).

We want to use modal logic to investigate what is provable in PA about $(\mathsf{V}_{\mathsf{PA}}, \rhd)$. Part of the answer is provided by the interpretability logic ILM. The language of ILM extends the basic modal language by having a binary modality $\rhd$. The intended interpretations of $\Box$ and $\rhd$ in ILM are formalized provability (in PA) and formalized interpretability (between finite extensions of PA) respectively. The axioms of ILM include K (for $\Box$), Löb's principle $\Box(\Box A \to A) \to \Box A$, and furthermore

(1) J1 $\Box(A \to B) \to A \rhd B$
(2) J2 $(A \rhd B) \wedge (B \rhd C) \to (A \rhd C)$
(3) J3 $(A \rhd C) \wedge (B \rhd C) \to (A \vee B) \rhd C$
(4) J4 $A \rhd B \to (\Diamond A \to \Diamond B)$

[1] An important example is provided by the well–known fact that Peano Arithmetic is interpretable in the Zermelo–Fraenkel set, via a translation that maps statements about natural numbers to statements about finite ordinals.

[2] The relation $\rhd$ between finite extensions of PA induces a partial order $\rhd$ on V_{PA}.

(5) J5 $\Diamond A \rhd A$
(6) M $A \rhd B \to (A \wedge \Box C) \rhd (B \wedge \Box C)$

The rules of ILM are modus ponens, and neccessitation for $\Box$. The fragment of ILM containing only $\Box$ is the provability logic GL. The arithmetical completeness of ILM was proven independently in [3] and in [4].

Applying the principles of ILM, it is not difficult to see that PA proves that $(\mathsf{V}_{\mathsf{PA}}, \rhd)$ is a lower semilattice. In particular, the infimum of the degrees of $\mathsf{PA}+A$ and $\mathsf{PA}+B$ is simply the degree of $\mathsf{PA}+(A \vee B)$. The existence of the supremum, on the other hand, can not be expressed in the language of ILM. A modal analysis of the supremum in $(\mathsf{V}_{\mathsf{PA}}, \rhd)$ apparently requires us to add to ILM a binary modal operator $\owedge$ for the supremum, together with the axiom:

$$(C \rhd A) \wedge (C \rhd B) \leftrightarrow C \rhd A \owedge B. \tag{1}$$

The intended meaning of the new modality $\owedge$ should then be an arithmetical formula $\sigma(x, y)$ with the property that for all sentences ϕ, ψ, and χ of the language of[3] PA,

$$\vdash_{\mathsf{PA}} (\chi \rhd \phi) \wedge (\chi \rhd \psi) \leftrightarrow \chi \rhd \sigma(\phi, \psi). \tag{2}$$

However, as was discovered by Volodya Shavrukov, we can do better than that, namely there exists, in PA, a supremum operator that is *unary*. By this we mean a formula $\sigma(x)$ such that for all sentences ϕ, ψ, and χ of the language of PA,

$$\vdash_{\mathsf{PA}} (\chi \rhd \phi) \wedge (\chi \rhd \psi) \leftrightarrow \chi \rhd \sigma(\phi) \wedge \sigma(\psi). \tag{3}$$

Thus we can add to ILM a new unary modality $\triangledown$ whose intended interpretation is $\sigma(x)$, and the corresponding axiom $(C \rhd A) \wedge (C \rhd B) \leftrightarrow C \rhd (\triangledown A \wedge \triangledown B)$. The rest of this note deals with a certain unary supremum operator. Section 2 establishes some of its properties as proven in PA, and Section 3 contains a modal analysis.

2 A Unary Supremum Operator

Consider a sequence of theories defined by[4] $T_0 = \mathsf{I}\Sigma_0 + \mathsf{exp}$ and $T_{n+1} = \mathsf{I}\Sigma_{n+1}$. Write $\vdash_n$ for provability in T_n, let $\Box_x$ stand for the provability predicate of T_x, and $\Box_x^{\Pi_1}$ for the provability predicate of T_x together with all true Π_1–sentences. As usual, we write $\Box$ for the provability predicate of PA, and $\Diamond\phi$ is defined as $\neg\Box\neg\phi$ (similarly $\Diamond_x$ and $\Diamond_x^{\Pi_1}$). The following facts are provable in PA, for any sentence ϕ and for any n (for more information about this particular stratification sequence, see [5]):

(1) $\vdash_{\mathsf{PA}} \phi \;\Leftrightarrow\; \exists n \vdash_n \phi$
(2) $\vdash_{n+1} \forall y\, (\Box_n^{\Pi_1}\phi(y) \to \phi(y))$, where $\phi(y)$ is Π_{n+3}

[3] We shall identify sentences of the language of arithmetic with their gödelnumbers.
[4] The theory $\mathsf{I}\Sigma_x$ is PA with induction restricted to Σ_x–formulas.

(3) $\vdash_n \phi \;\Rightarrow\; \vdash_{n+1} \phi$

The intuition is that we have a growing (iii) hierarchy of theories that is a stratification of PA (i), each level being sufficiently stronger than the previous one (ii). Define $\triangledown A := \forall x\,(\Diamond^{\Pi_1}_{x-1}\top \to \Diamond_x A)$. The following theorem states that $\triangledown$ is a unary supremum operator for the lattice $(\mathsf{V}_{\mathsf{PA}}, \rhd)$.

Theorem 1. *Let ϕ, ψ and χ be sentences of the language of* PA*. Then*

$$\vdash_{\mathsf{PA}} (\chi \rhd \phi) \wedge (\chi \rhd \psi) \leftrightarrow \chi \rhd (\triangledown\phi \wedge \triangledown\psi). \tag{4}$$

Proof. We use the formalized version of the Orey-Hájek characterisation of interpretability (see [6]), whereby $\phi \rhd \psi$ if and only if $\phi \vdash_{\mathsf{PA}} \Diamond_n\psi$ for all n. Thus it suffices to show $\triangledown\phi \vdash \Diamond_n\phi$ and $\Diamond_n\phi \wedge \Diamond_n\psi \vdash_{\mathsf{PA}} \Diamond_n(\triangledown\phi \wedge \triangledown\psi)$ for all n. The first follows by properties of our chosen stratification sequence $\{T_n\}_{n\in\omega}$ listed above, for the latter we also use that GL is valid for $\Box_x$ (in PA) for all x.

Let us think of the sentence $\triangledown\phi$ as a consistency statement for $\mathsf{PA} + \phi$. This perspective turns out to be rather useful. The corresponding notion of provability is then given by $\triangle\phi := \neg\triangledown\neg\phi$. Note that $\triangle\phi$ is the sentence $\exists x\,(\Box_x\phi \wedge \Diamond^{\Pi_1}_{x-1}\top)$, and thus $\triangle$ embodies a stronger notion of provability than the usual $\Box$. If ϕ is $\triangle$–provable, then ϕ is provable in some $\mathsf{I}\Sigma_x$ (thus provable in the usual sense), and furthermore the theory $\mathsf{I}\Sigma_{x-1}$ together with all true Π_1–sentences is consistent[5]. Nevertheless, $\triangle$ turns out to be a well-behaved provability predicate, in the sense that it obeys the principles of GL.

Theorem 2. *Let ϕ be a sentence of the language of arithmetic. Then*

(1) $\vdash_{\mathsf{PA}} \phi \;\Rightarrow\; \vdash_{\mathsf{PA}} \triangle\phi$
(2) $\vdash_{\mathsf{PA}} \triangle(\phi \to \psi) \to (\triangle\phi \to \triangle\psi)$
(3) $\vdash_{\mathsf{PA}} \triangle\phi \to \triangle\triangle\phi$

Applying the Fixed Point Lemma, we can use the usual argument to establish that also Löb's principle is valid for $\triangle$, i.e. that $\vdash_{\mathsf{PA}} \triangle(\triangle\phi \to \phi) \to \triangle\phi$. It follows that the modal system GL is sound for $\triangle$ in PA. To end this section, we list some principles concerning the interaction of $\Box$ and $\triangle$.

Theorem 3.

i.	$\vdash_{\mathsf{PA}} \triangle\phi \to \Box\phi$	(T1)
ii.	$\vdash_{\mathsf{PA}} \Box\phi \to \triangle\Box\phi$	(T2)
iii.	$\vdash_{\mathsf{PA}} \Box\phi \to \Box\triangle\phi$	(T3)
iv.	$\vdash_{\mathsf{PA}} \Box\triangle\phi \to \Box\phi$	(T4)

[5] Thus $\triangle$ is similar to the Feferman provability predicate $\Box^{\mathsf{F}}$ given by

$$\Box^{\mathsf{F}}\phi := \exists x\,(\Box_x\phi \wedge \Diamond_x\top).$$

Proof. Note that $\vdash_{\mathsf{PA}} (\phi \rhd \triangledown\phi) \wedge (\triangledown\phi \rhd \phi)$ by Theorem 1, and so by axiom $\mathsf{J4}$ of ILM, $\vdash_{\mathsf{PA}} (\Diamond\phi \to \Diamond\triangledown\phi) \wedge (\Diamond\triangledown\phi \to \Diamond\phi)$. We obtain $\mathsf{T3}$ and $\mathsf{T4}$ by contraposition. $\mathsf{T1}$ is clear by the definition of $\triangle$, and for $\mathsf{T2}$ we use that $\vdash_{\mathsf{PA}} \psi \to \Box_1\psi$ for any $\psi \in \Sigma_1$.

3 A Bimodal Provability Logic

We consider the joint provability logic GLT of $\Box$ and $\triangle$. The system GLT has as axioms the GL axioms for $\Box$ and $\triangle$, and the principles $\mathsf{T1}$–$\mathsf{T4}$ from Theorem 3 above. The rules of GLT are modus ponens, and necessitation for $\Box$ and $\triangle$. Arithmetical soundness of GLT follows by theorems 2 and 3 above.

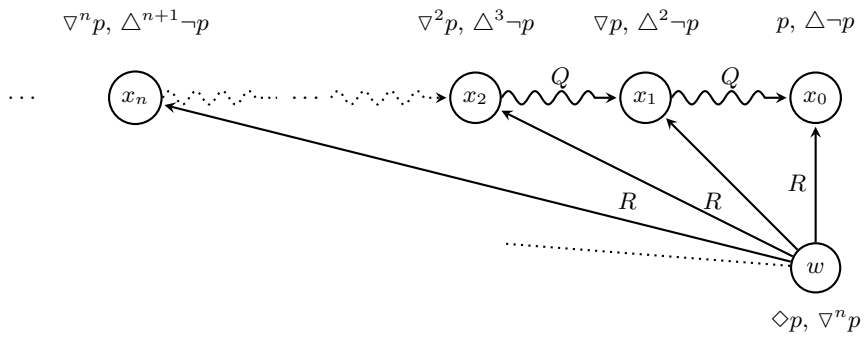
Definition 1. *A GLT frame is a triple $\langle W, R, Q\rangle$ with $W \neq \emptyset$, both R and Q conversely well–founded and transitive, $R \subseteq Q$, $Q \circ R \subseteq R$, $R \circ Q \subseteq R$, and $R \subseteq R \circ Q$. A GLT–model is a quadruple $\langle W, R, Q, \Vdash\rangle$, where $\langle W, R, Q\rangle$ is a GLT–frame and $\Vdash$ is a forcing relation on $\langle W, R, Q\rangle$ satisfying the usual clauses with R and Q as the accessibility relations for $\Box$ and $\triangle$ respectively.*

Due to the conversely well–foundedness of R and Q, the condition $R \subseteq R \circ Q$ can be equivalently formulated as: if wRx_0, then there is a sequence $\{x_n\}_{n\in\omega}$ of distinct nodes with wRx_n and $x_{n+1}Qx_n$ for all n. A GLT–frame with at least one R–relation is therefore necessarily infinite.

Theorem 4 (Modal Completeness of GLT). *$\vdash_{\mathsf{GLT}} A$ if and only if A is valid on all GLT–frames.*

Proof. The non–trivial direction is to find, given a formula A with $\nvdash_{\mathsf{GLT}} A$, a GLT–model where A is not true at a node. We use the construction method.

Figure 1: A GLT–model with $w \Vdash \Diamond p$



Whenever a node w (a maximal consistent set) contains a formula of the form $\Diamond B$, we add an infinite sequence $\{x_n\}_{n\in\omega}$ of nodes to the frame, with $x_0 \Vdash B$. To ensure the conversely well–foundedness of R and Q, each of the new R and Q relations needs to be witnessed by a unique $\triangle$–formula contained in some set $\mathcal{D}$ with limited (though possibly infinite) size. We take as $\mathcal{D}$ the set $\{\triangle^n\neg B \mid n \in \omega\}$ (where $\triangle^n A$ is defined inductively by letting $\triangle^0 A := A$ and $\triangle^{n+1} A := \triangle\triangle^n A$). The situation is illustrated in Figure 1, with $B = p$ (the relations resulting from frame conditions of GLT, e.g. transitivity of Q, are omitted).

Note that since GLT–frames are in general infinite, Theorem 4 does not have the decidability of GLT as a consequence.

4 Lindström's Bimodal System for Parikh Provability

As it turns out, the modal system GLT has been studied by Per Lindström already in 1994 ([7]). Interestingly, the arithmetical interpretation differs from our case. Namely, $\triangle$ is interpreted as the usual provability predicate of PA, and $\Box$ is interpreted as the provability predicate of the system PA together with Parikh's rule:

$$\text{from } \mathsf{Pr}_{\mathsf{PA}}\phi, \text{ infer } \phi \tag{5}$$

While adding Parikh's rule does not lead to new theorems (it is admissible in PA), it does lead to shorter proofs. Lindström proves arithmetical soundness and completeness of GLT with respect to this interpretation of the modalities.

Lindtsröm also proves modal completeness of GLT. Also here the semantics is different from our case. Namely, GLT is proven to be modally sound and complete with respect to Kripke frames $\langle W, R, \Vdash\rangle$, where R is transitive and conversely well–founded, and the semantics of $\Box$ is defined in the following way:

$$w \Vdash \Box A :\Leftrightarrow x \Vdash A \text{ for all } x \text{ such that } \{z \mid wRz \wedge zRx\} \text{ is infinite.} \tag{6}$$

It is easy to see that all the axioms of GLT are valid under this interpretation of the $\Box$. Lindström also proves modal completeness of GLT with respect to a simpler class of frames, obtaining decidability as a corollary.

5 Future Work

It remains to be explored whether the fact that GLT has two meaningful arithmetical interpretations – one discussed in this note, and the other one studied by Lindström – is more than a coincidence. Independent from that, however, we intend to prove arithmetical completeness of the modal system GLT. The final goal is, of course, to add the binary modality $\rhd$ of ILM to GLT, and prove arithmetical completeness for the resulting system.

Acknowledgements. The idea of unary supremum operators is due to Volodya Shavrukov; he was also the first one to give an example of such an operator. The operator $\triangledown$ was obtained by analysing the latter. I thank Albert Visser for providing the expertise needed for the arithmetical matters contained in this paper. As for the modal side, I am grateful for the assistance given by Dick de Jongh and Frank Veltman.

References

1. Tarski, A. and Mostowski, A. and Robinson, R.M.: Undecidable theories. North–Holland, Amsterdam, 1953.
2. Švejdar, V.: Degrees of interpretability. Commentationes Mathematicae Universitatis Carolinae, vol 19, pp.789–813, 1978.
3. Shavrukov, V.Yu.: The logic of relative interpretability over Peano arithmetic (in Russian). Steklov Mathematical Institute, Moscow, No.5, 1988.
4. Berarducci, A.: The interpretability logic of Peano arithmetic. The Journal of Symbolic Logic, vol 55, pp.1059–1098, 1990.
5. Shavrukov, V.Yu and Visser. A: Uniform Density in Lindenbaum Algebras. Manuscript. 2012.
6. Joosten, J.J.: Interpretability Formalised. PhD Thesis, University of Utrecht, 2004.
7. Lindström, P.: On Parikh Provability - an Exercise in Modal Logic. Modality Matters: Twenty-Five Essays in Honour of Krister Segerberg. Henrik Lagerlund and Sten Lindströ and Rysiek Sliwinski (eds). Uppsala Philosophical Studies 53, pp.279–288, 2006.

Towards a Propositional Logic for Reversible Logic Circuits

Robin Kaarsgaard

DIKU, Department of Computer Science, University of Copenhagen
robin@diku.dk

Abstract Established theories of reversible circuit logic have so far constrained themselves only to the semantic level, unable to reap the potential benefits of reasoning about such circuits offered by a deductive system complete with respect to these semantics. This paper details the development of such a deductive system, based on Toffoli's theory of reversible computing: A syntactic representation of key parts of reversible circuits is developed, as is a propositional logic based on these primitives, and its key metatheorems are derived.

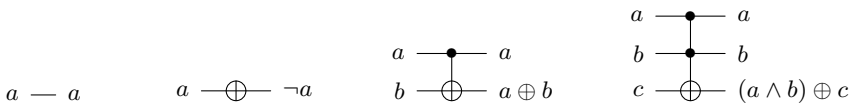
1 Introduction

Irreversible computing comes at a minimum cost, paid in energy dissipated as heat across the computing process – this was first argued by Landauer in 1961 [1] and has more recently been verified experimentally by Bérut et al. [2]. This intimate relationship between information preservation and thermodynamic reversibility has led to the development of theories of reversible digital logic circuits, which circumvent this lower limit of energy consumption by ensuring logical reversibility.

While several theories of reversible circuit logic exist – most notably those of Toffoli and Fredkin [3][4] – none of these have, until now, been explored as formal deductive systems, opting instead to focus only on the semantic side of the coin. Though this purely semantic approach is sufficient for deriving key results such as universality, it leaves something to be desired with respect to, e.g., the optimization of such circuits as sound rewriting rules can only be developed by painstaking computation of truth tables.

This article seeks to explore reversible circuit logic from the other side of the coin, through the development of a formal propositional logic that is sound and complete with respect to the semantics already developed by Toffoli. In particular, this yields a provable equivalence relation strong enough to perform line-by-line optimization of reversible circuits, though it fails to capture the structural properties of such circuits.

Previous work has gone into investigating reversibility in logic, in particular with Sparks & Sabry's reversible logic of RL [5]. However, the goal of RL is fundamentally different from ours, though it belongs in the same intersection between logic and computation, namely to explore the nature of reversible computing by developing a deductive system which has the reversibility of *proofs* as



(a) ID gate. (b) NOT gate. (c) FEYNMAN gate. (d) TOFFOLI gate.

Figure 1: The operational definitions of ID, NOT, FEYNMAN and TOFFOLI gates. The operation $- \oplus -$ denotes exclusive disjunction.

an intrinsic property. On the other hand, our goal is to construct a deductive system specifically in correspondence with Toffoli's theory of reversible circuit logic – a proof theoretic notion of reversibility is neither intended nor relevant, so our logic and RL are complementary rather than in competition.

This article is structured as follows: Section 2 gives a brief introduction to Toffoli's reversible gate set, and develops a syntax and semantics for the representation of (the most relevant parts of) such circuits as logical formulae; Section 3 presents the propositional logic and its metatheorems; Section 4 details the possible applications of this logic; and Section 5 concludes on the presented results, and discusses avenues for future research.

2 Representing reversible logic circuits

The theory developed bases itself on Toffoli's reversible gate set [3], consisting of the ID and NOT gates (semantically identical to their counterparts in traditional circuit logic, as these are reversible by definition) as well as the FEYNMAN and TOFFOLI gates, sometimes also called "controlled NOT" respectively "controlled-controlled NOT" (see Figure 1). Further, we require the existence of lines of constant value 0 and 1, often called *local storage* or *ancillae* lines. As in traditional circuit logic, these gates and lines may be composed both horizontally (as in ordinary function composition) as well as vertically (by computation in parallel) as long as no loops are formed. In contrast to traditional circuit logic, however, fan-out is not allowed (as this would violate reversibility). In addition, the control lines (marked with black dots) may be placed anywhere relative to the lines they control; see, e.g., the circuit in Figure 2.

From the operational definition of these gates, it seems that one can adequately model the target line of such gates through a propositional logic that

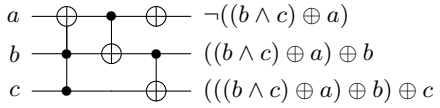


Figure 2: A reversible circuit annotated according to the operational definitions of its constituent gates.

A	$\neg A$
0	1
1	0

(a) Negation.

γ	A	$\gamma \bullet\!\!- A$
0	0	1
0	1	0
1	0	0
1	1	1

(b) Control.

γ	A	$\gamma \wedge A$
0	0	0
0	1	0
1	0	0
1	1	1

(c) Conjunction.

$\bot$
0

(d) Bottom.

$\top$
1

(e) Top.

Figure 3: Truth tables for propositions and conjunctions.

restricts itself to propositions that are either negations, exclusive disjunctions, or conjunctions guarded by an exclusive disjunction. While this is certainly possible, working directly with exclusive disjunction is awkward, and does little to explain how, e.g., the Feynman gate can be thought of as a "controlled NOT" gate. Indeed, if the Feynman gate is "controlled NOT", what is control?

The answer is plain and simple bi-implication; however, to avoid confusion and stress that it is an atomic propositional form, we will denote it with a distinguished symbol, and write $a \bullet\!\!- b$ for "a control b". In this way, we can rewrite the target line of the FEYNMAN resp. TOFFOLI gates in Figure 1 to $a \bullet\!\!- \neg b$ resp. $(a \wedge b) \bullet\!\!- \neg c$. This turns out to give pleasant rules for forming and decomposing the representation of such gates, but comes at the price that we need to be able to produce multiple conclusions from the same context, as we will not allow uncontrolled conjunction, since these do not correspond directly to the target line of any gate in the gate set. The immediate correspondence between target lines of reversible logic gates and propositions is central in the design, in fact its *raison d'être*; if we allow uncontrolled conjunctions as first class propositions, we lose this property.

For this reason, we introduce formulae as (nonempty) lists of propositions; non-singleton formulae are not themselves propositions (in that one cannot talk about the truth or falsity of a formula as a whole) – however, one *can*, at least at the meta-level, reasonably talk about the truth of some or all of a formula's constituent propositions. All in all, this gives us the following syntax (where a denotes any propositional atom):

$$\begin{aligned}
A, B, C &:= a \mid \neg A \mid \gamma \bullet\!\!- A \mid \bot \mid \top && \text{(Propositions)} \\
\gamma &:= A \mid \gamma \wedge A && \text{(Conjunctions)} \\
\Phi, \Psi, \Pi &:= A \mid \Phi, A && \text{(Formulae)} \\
\Gamma, \Delta, \Theta &:= \cdot \mid \Gamma, A && \text{(Contexts)}
\end{aligned}$$

In addition to this, we shall need an auxiliary function for turning formulae into conjunctions – we call this the *conjunctive lift* of a formula Φ, denote it $\lceil \Phi \rceil$, and define it in the obvious way:

$$\lceil \Phi \rceil = \begin{cases} A & \text{if } \Phi = A \\ \lceil \Psi \rceil \wedge A & \text{if } \Phi = \Psi, A \end{cases}$$

Core rules

$$\frac{}{\Gamma, A \vdash A}\ (\textsc{Id}) \qquad \frac{\Gamma \vdash A \qquad \Gamma, A \vdash B}{\Gamma \vdash B}\ (\textsc{Cut})$$

Structural rules

$$\frac{\Gamma \vdash \Phi}{\Gamma, \Delta \vdash \Phi}\ (\textsc{Wkn}) \qquad \frac{\Gamma \vdash \Phi \qquad \Gamma \vdash \Psi}{\Gamma \vdash \Phi, \Psi}\ (\textsc{Cnc})$$

$$\frac{\Gamma, \Delta, \Delta \vdash \Phi}{\Gamma, \Delta \vdash \Phi}\ (\textsc{Cnt}) \qquad \frac{\Gamma, A, \Delta, B, \Theta \vdash \Phi}{\Gamma, B, \Delta, A, \Theta \vdash \Phi}\ (\textsc{Exc})$$

Logical rules

$$\frac{\Gamma \vdash \Phi \qquad \Gamma, A \vdash \Psi}{\Gamma, \lceil \Phi \rceil \bullet\!\!- A \vdash \Psi}\ (\bullet\!\!- \mathrm{L}_1) \qquad \frac{\Gamma, \Phi \vdash A \qquad \Gamma, A \vdash \Phi}{\Gamma \vdash \lceil \Phi \rceil \bullet\!\!- A}\ (\bullet\!\!- \mathrm{R})$$

$$\frac{\Gamma \vdash A \qquad \Gamma, \Phi \vdash \Psi}{\Gamma, \lceil \Phi \rceil \bullet\!\!- A \vdash \Psi}\ (\bullet\!\!- \mathrm{L}_2)$$

$$\frac{\Gamma \vdash A \qquad \Gamma, \bot \vdash \Phi}{\Gamma, \neg A \vdash \Phi}\ (\neg\mathrm{L}) \qquad \frac{\Gamma, A \vdash \bot}{\Gamma \vdash \neg A}\ (\neg\mathrm{R})$$

$$\frac{}{\Gamma, \bot \vdash A}\ (\bot\mathrm{L}) \qquad \text{(no right rule for } \bot\text{)}$$

$$\text{(no left rule for } \top\text{)} \qquad \frac{}{\Gamma \vdash \top}\ (\top\mathrm{R})$$

Classical rules

$$\frac{\Gamma, A \vdash B \qquad \Gamma, \neg A \vdash B}{\Gamma \vdash B}\ (\textsc{Lem}) \qquad \frac{\Gamma, A \vdash \Phi}{\Gamma, \neg\neg A \vdash \Phi}\ (\neg\neg\mathrm{L})$$

Figure 4: The inference rules of LRS.

Classical, bivalent truth table semantics are used to assign truth values to propositions according to the tables in Figure 3. Note that even though uncontrolled conjunctions are not syntactically recognized as propositions, since the truth value of a control may depend on the truth value of a conjunction, we need to assign truth values to conjunctions in order to ensure that all propositions can be assigned truth values.

3 A classical logic of reversible structures

The deductive system, which we shall call the *logic of reversible structures* (or LRS for short, see Figure 4), is a sequent calculus, inspired in its design by Gentzen's famous calculi LJ and LK [6]. It has a single judgement form, $\Gamma \vdash \Phi$, taken to mean that Γ proves all of the propositions of Φ, and a corresponding entailment relation $\Gamma \vDash \Phi$ taken to mean that all propositions of Φ are true

whenever all propositions of Γ are true (as specified by the truth tables in the previous section).

Following the structure of Gentzen's calculi, LRS bases itself on identity and cut as core rules, has weakening, contraction, and exchange as its structural rules, and separates its logical rules into groups of left respectively right rules, corresponding (somewhat) to the elimination respectively introduction rules in natural deduction style. The only things out of the ordinary in the structural rules can be found in the concatenation rule (Cnc) for forming formulae, as well as the fact that the Law of the Excluded Middle takes a more direct form as an explicit case analysis, due to the omission of disjunction as a syntactic form for propositions.

Since we expect control to behave as an "atomic bi-implication", its logical rules must reflect this: Thus, we form such propositions by assuming one side of the control in the context and proving the other (and vice versa), and decompose them by proving that one side of the control holds in the context, justifying that one may assume the other side in the context for future derivations. This is also where the conjunctive lift comes into play, as it allows for a smooth composition and decomposition of the controlled conjunctions into the context.

In Section 4 (Figure 5) we give an example of a useful derivation.

3.1 Soundness, completeness and universality

Similar to the proof of soundness for classical natural deduction, the soundness proof of LRS with respect to its semantics is a mostly mechanical exercise which can be skipped on a first reading.

Theorem 1 (Soundness). *If $\Gamma \vdash \Phi$ then $\Gamma \vDash \Phi$.*

Proof. By induction on the structure of the derivation of $\Gamma \vdash \Phi$.

While the proof of soundness is straightforward, proving completeness is somewhat more involved, as the absense of ordinary implication obscures whether an analogue to the deduction theorem of propositional logic exists.

Instead we prove something stronger, namely the universality of LRS with respect to classical propositional logic, here in the form of classical natural deduction (ND). In short, what this means is that every proposition in ND has an equivalent encoding in LRS, and that a proof of the original proposition in ND may be transformed into a proof of the encoded proposition in LRS. In its semantic form, this relationship goes back to Toffoli's original article, stating that every (not necessarily reversible) boolean function may be computed by a reversible circuit, if one allow *local storage* lines (input lines of constant value) and *garbage outputs* (superfluous output values not part of the original function) to occur [3, Section 4].

To avoid confusion between judgments and entailments in ND and LRS, we shall denote such in ND by specific addition of subscript ND to the turnstile. The semantics-preserving encoding of ND propositions and contexts is given by the following definition.

Definition 1. *Let φ be a proposition in ND. The semantics-preserving encoding of φ in LRS, denoted $\widehat{\varphi}$, is defined as follows by induction on the structure of φ:*

$$\widehat{\varphi} = \begin{cases} a & \text{if } \varphi = a \\ \top & \text{if } \varphi = \top \\ \bot & \text{if } \varphi = \bot \\ \neg\widehat{A} & \text{if } \varphi = \neg A \\ (\widehat{A} \wedge \widehat{B}) \bullet\!\!- \top & \text{if } \varphi = A \wedge B \\ (\neg\widehat{A} \wedge \neg\widehat{B}) \bullet\!\!- \bot & \text{if } \varphi = A \vee B \\ (\widehat{A} \wedge \neg\widehat{B}) \bullet\!\!- \bot & \text{if } \varphi = A \to B \end{cases}$$

Further, we define this encoding on contexts by letting it distribute over the comma. That this encoding is sound follows by nested induction on Γ and φ by drawing up truth tables. With this in place, we have all the tools we need to show the universality of LRS with respect to ND.

Theorem 2 (Universality). *If $\Gamma \vdash_{ND} \varphi$ then $\widehat{\Gamma} \vdash \widehat{\varphi}$.*

Proof. By induction on the structure of the derivation of $\Gamma \vdash_{ND} \varphi$.

The proof goes smoothly with regards to the structural rules of ND as well as the logical rules for units, negations and conjunctions, while the real work lies in the rules for disjunction and implication. Since the encoding of disjunction and implication relies on classical reasoning, use of the Lem rule is required in some of these cases for the proof to go through.

Since classical natural deduction is complete with respect its truth table semantics [7], we can use the universality theorem to prove completeness of LRS, by providing a semantics preserving decoding of LRS propositions into ND. In addition, this requires a few technical lemmas to go through. The decoding is defined analogously to Definition 1, and its soundness follows straightforwardly by nested induction on Γ and φ, drawing up truth tables in each step.

The gist of the completeness proof is as follows: If we can show that the encoding of the decoding of a proposition in LRS is provably equivalent to the original proposition, we can take a true proposition of LRS, decode it into ND, and use completeness of ND to get a proof. Feeding this proof through universality then gives us a proof in LRS of the encoding of the decoding of the original proposition; but since we have a lemma stating that these are provably equivalent, we can simply use the Cut rule to inductively substitute it in.

Theorem 3 (Completeness). *If $\Gamma \vDash \Phi$ then $\Gamma \vdash \Phi$.*

Proof. By nested induction on the structure of Γ and Φ.

4 Applications

An obvious application of LRS – in fact, the intended purpose of it – is in rewriting and optimization of reversible logic circuits, in particular in determining

$$\dfrac{\dfrac{\dfrac{}{\neg B \vdash \neg B}\,(\textsc{Id})}{\neg B, \neg A \vdash \neg B}\,(\textsc{Wkn}) \qquad \dfrac{\dfrac{\dfrac{}{\neg B, A \vdash A}\,(\textsc{Id}) \qquad \dfrac{}{\neg B, A, \bot \vdash \bot}\,(\textsc{Id})}{\neg B, A, \neg A \vdash \bot}\,(\neg\mathrm{L})}{\neg B, \neg A, A \vdash \bot}\,(\textsc{Exc})}{\dfrac{\neg B, \neg A, A \bullet\!\!- \neg B \vdash \bot}{\neg B, \neg A \vdash \neg(A \bullet\!\!- \neg B)}\,(\neg\mathrm{R})}\,(\bullet\!\!-\mathrm{L}_2)$$

(a) Derivation of $\mathcal{D}_1$.

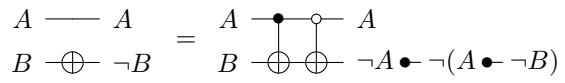
$$\dfrac{\dfrac{\dfrac{\dfrac{}{A, A, \neg B \vdash \neg B}}{\neg B, A, A \vdash \neg B}\,(\textsc{Exc}) \qquad \dfrac{\dfrac{\dfrac{}{\neg B, \neg B, A \vdash A}\,(\textsc{Id})}{\neg B, A, \neg B \vdash A}\,(\textsc{Exc})}{\neg B, A \vdash A \bullet\!\!- \neg B}\,(\bullet\!\!-\mathrm{R}) \qquad \dfrac{}{\neg B, A, \bot \vdash \bot}\,(\textsc{Id})}{\dfrac{\dfrac{\neg B, A, \neg(A \bullet\!\!- \neg B) \vdash \bot}{\neg B, \neg(A \bullet\!\!- \neg B), A \vdash \bot}\,(\textsc{Exc})}{\neg B, \neg(A \bullet\!\!- \neg B) \vdash \neg A}\,(\neg\mathrm{R})}\,(\neg\mathrm{L})$$

(b) Derivation of $\mathcal{D}_2$.

$$\dfrac{\begin{array}{cc}\mathcal{D}_1 & \mathcal{D}_2 \\ \neg B, \neg A \vdash \neg(A \bullet\!\!- \neg B) & \neg B, \neg(A \bullet\!\!- \neg B) \vdash \neg A\end{array}}{\neg B \vdash \neg A \bullet\!\!- \neg(A \bullet\!\!- \neg B)}\,(\bullet\!\!-\mathrm{R})$$

Figure 5: Derivation of $\neg B \vdash \neg A \bullet\!\!- \neg(A \bullet\!\!- \neg B)$.

the completeness of such systems. A particularly interesting rewriting system is that of Soeken & Thomsen [8], as many of their rewriting rules act on only one line at a time, enabling their representation in LRS. One such rule is Soeken & Thomsen's R3, which in the first case, fully annotated, states that



This rule is a theorem of LRS, i.e. we have $\neg B \dashv\!\vdash \neg A \bullet\!\!- \neg(A \bullet\!\!- \neg B)$ – the left-to-right direction is shown in Figure 5, and the right-to-left direction is also derivable, though it requires appeal to the Law of the Excluded Middle.

Questions of completeness are not quite ready to be answered by LRS, however, due to its inability to capture the independence of computation in parallel. Naïvely, one could assume that the juxtaposition of propositions offered by formulae would be enough to capture the behaviour of multiple lines simultaneously in a reversible circuit – and, indeed, this was the initial intention. However, this approach fails to capture both the order of such circuits (since the exchange rule allows arbitrary permutation to occur on both sides of the turnstile), and the fact that the truth of one proposition may depend on another; for example, $A, A \bullet\!\!- B \dashv\!\vdash A, B$ even though they are semantically different circuits (when erroneously interpreting juxtaposition as computation in parallel) – not because

the logic is unsound, but because the linearly additive nature of the concatenation rule fails to capture what we really mean, namely that $A \bullet\!\!- B \dashv\!\vdash B$ should hold independently of the truth of A, which is not the case.

5 Conclusion and future work

The author believes that LRS has succeeded in capturing essential properties of reversible circuit logic, in particular its relation to traditional, irreversible circuit logic through the universality theorem. Though the logic is not yet able to faithfully capture the entire structure of reversible circuits, and in this way provide a provable equivalence relation with meaning akin to an equivalence of reversible circuits, it is the opinion of the author that it serves as a useful foundation for further studies in the equality of reversible circuits.

Since LRS fails to capture the structural aspects of reversible circuit logic, most notably in its inability to preserve line order and handle parallel lines independently of one another, it seems prudent to ask what will. It is conjectured that an ordered and linear version of the logic may answer this, as ordered multiplicative conjunction both enforces independence of parallel lines *and* ensures that arbitrary line permutations cannot occur. This is a current research topic, however, with only very preliminary results thus far.

Acknowledgements

The author wishes to thank (in no particular order) Holger Bock Axelsen, Michael Kirkedal Thomsen, and Robert Glück for their tremendous encouragement, insights, and helpful comments, and the three anonymous reviewers for their valuable feedback.

References

1. Landauer, R.: Irreversibility and heat generation in the computing process. IBM Journal of Research and Development **5**(3) (1961) 261–269
2. Bérut, A., Arakelyan, A., Petrosyan, A., Ciliberto, S., Dillenschneider, R., Lutz, E.: Experimental verification of Landauer's principle linking information and thermodynamics. Nature **483**(7388) (2012) 187–189
3. Toffoli, T.: Reversible computing. In: Proceedings of the Colloquium on Automata, Languages and Programming, Springer-Verlag (1980) 632–644
4. Fredkin, E., Toffoli, T.: Conservative logic. International Journal of Theoretical Physics **21**(3-4) (1982) 219–253
5. Sparks, Z., Sabry, A.: Superstructural reversible logic. Unpublished, fetched from `https://www.cs.indiana.edu/~sabry/papers/reversible-logic.pdf` (2013)
6. Gentzen, G.: Untersuchungen über das logische Schließen. Mathematische Zeitschrift **39**(1) (1935) 176–210
7. Huth, M., Ryan, M.: Logic in computer science: Modelling and reasoning about systems. Cambridge University Press (2004)
8. Soeken, M., Thomsen, M.K.: White dots do matter: Rewriting reversible logic circuits. In Dueck, G.W., Miller, D.M., eds.: Reversible Computation. Volume 7948 of LNCS. (2013) 196–208

A Universal Diagonal Schema by Fixed-Points

Ahmad Karimi

Department of Mathematics, Tarbiat Modares University, Tehran, IRAN.
ahmad.m.karimi@gmail.com

Abstract A universal schema for diagonalization was popularized by Yanofsky (2003) in which the existence of a (diagonolized–out and contradictory) object implies the existence of a fixed–point for a certain function. It was shown that many self–referential paradoxes and diagonally proved theorems can fit in that schema. Here, we fit more theorems in the universal schema of diagonalization, like some new proofs of Boolos (1997) for Cantor's theorem on the non–equinumerosity of a set with its powerset. Also it is shown that Priest's (1997) inclosure schema can fit in our universal diagonal/fixed–point schema. Furthermore, we formalize a reading of Yablo's paradox, the most challenging paradox in the recent years, in the framework of Linear Temporal Logic (LTL) and the diagonal schema, and show how Yablo's paradox involves circularity by presenting it in the framework of LTL. Indeed, we turn Yablo's paradox into a genuine mathematico logical theorem. This is the first time that Yablo's paradox becomes a (new) theorem in mathematics and logic.

Keywords: Diagonalization, Self–Reference, Fixed–Points, Cantor's Theorem, Yablo's Paradox, Linear Temporal Logic.

1 Introduction

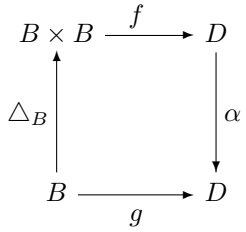
In 1906, Russell [13] showed that all the known set-theoretic paradoxes (till then) had a common form. In 1969, Lawvere [9] used the language of category theory to achieve a deeper unification, embracing not only the set-theoretic paradoxes but incompleteness phenomena as well. To be precise, Lawvere gave a common form to Cantor's theorem about power sets, Russell's paradox, Tarski's theorem on the undefinability of truth, and Gödel's first incompleteness theorem. In 2003, Yanofsky [16] extended Lawvere's ideas using straightforward set-theoretic language and proposed a universal schema for diagonalization based on Cantor's theorem. In this universal schema for diagonalization, the existence of a certain (diagonalized-out and contradictory) object implies the existence of a fixed-point for a certain function. He showed how self-referential paradoxes, incompleteness, and fixed-point theorems all emerge from the single generalized form of Cantor's theorem. Yanofsky extended Lawvere's analysis to include the Liar paradox, the paradoxes of Grelling and Richard, Turing's halting problem, an oracle version of the P=?NP problem, time travel paradoxes, Parikh sentences, Löb's Paradox and Rice's theorem.

In this paper, we fit more theorems in the universal schema of diagonalization, like some new proofs of Boolos [1] for Cantor's theorem on the non-equinumerosity of a set with its powerset. Furthermore, we formalize a reading of Yablo's paradox [15], the most challenging paradox in the recent years, in the framework of Linear Temporal Logic (LTL [8]) and the diagonal schema, and show how Yablo's paradox involves circularity by presenting it in the framework of LTL. Indeed, we turn Yablo's paradox into a genuine mathematico logical theorem. This is the first time that Yablo's paradox becomes a (new) theorem in mathematics and logic. We also show that Priest's [11] inclosure schema can fit in our universal diagonal/fixed-point schema. The inclosure schema was used by Priest for arguing for the self-referentiality of Yablo's sequence of sentences, in which no sentence directly refers to itself but the whole sequence does so. In the rest of the introduction we fix our notation and introduce the common framework.

1.1 Cantor's Theorem by Fixed–Points

Theorem 1 (Cantor). *Assume the function $\alpha : D \longrightarrow D$, for a set D, does not have any fixed point (i.e., $\alpha(d) \neq d$ for all $d \in D$). Then for any set B and any function $f : B \times B \to D$ there exists a function $g : B \to D$ that is not representable by f (i.e., for all $b \in B$, $g(-) \neq f(-, b)$).*

Proof. The desired function $g : x \mapsto \alpha(f(x, x))$ can be constructed as follows:



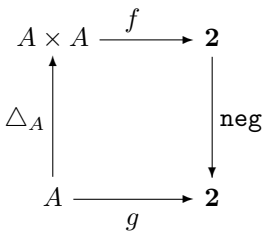
where $\triangle_B$ is the diagonal function of B ($\triangle_B(x) = \langle x, x\rangle$). If g is representable by f at $b \in B$, then $g(x) = f(x, b)$ for any $x \in B$, and in particular $g(b) = f(b, b)$. On the other hand by the definition of g we have $g(x) = \alpha(f(x, x))$ and particularly $g(b) = \alpha(f(b, b))$. It follows that $f(b, b)$ is a fixed–point of α; contradiction. Whence, the function g is not representable by f (at any $b \in B$). ⊠

For any set A we have $\mathscr{P}(A) \cong \mathbf{2}^A$ where $\mathbf{2} = \{0, 1\}$ and $\mathbf{2}^A$ is the set of all functions from A to $\mathbf{2}$. So, Cantor's theorem is equivalent to the non–existence of a surjection $A \to \mathbf{2}^A$. Putting it another way, Cantor's theorem says that for any function $f : A \times A \to \mathbf{2}$ there exists a function $g : A \to \mathbf{2}$ which is not representable by f (at any member of A). In this new setting, Cantor's proof goes as follows: let $\triangle_A : A \to A \times A$ be the diagonal function of A ($\triangle_A(x) = \langle x, x\rangle$) and let $\alpha : \mathbf{2} \to \mathbf{2}$ be a fixed function. Define $g : A \to \mathbf{2}$ by $g(x) = \alpha(f(\triangle_A(x)))$. If g is representable by f and fixed $a \in A$ then $f(a, a) = g(a) = \alpha(f(a, a))$, which shows that α has a fixed–point (namely, $f(a, a)$). So, for reaching to a

contradiction, we need to take a function $\alpha : \mathbf{2} \to \mathbf{2}$ which does not have any fixed–point; and the only such function (without any fixed–point) is the negation function $\mathtt{neg} : \mathbf{2} \to \mathbf{2}$, $\mathtt{neg}(i) = 1 - i$ for $i = 0, 1$. For a function $F : A \to \mathscr{P}(A)$ let $f : A \times A \to \mathbf{2}$ be defined as

$$f(a, a') = \begin{cases} 1 & \text{if } a \in F(a') \\ 0 & \text{if } a \notin F(a') \end{cases}$$

The function g constructed by the diagram (Yanofsky's framework)



is the characteristic function of the set $D = \{x \in A \mid x \notin F(x)\}$. That g is not representable by f (at any $a \in A$) is equivalent to saying that the set D is not in the range of F (i.e., $D \neq F(a)$ for any $a \in A$).

In the rest of the paper we will fit some theorems in the diagram of Yanofsky's framework by varying the set A (and the functions f).

2 Some Other Proofs for Cantor's Theorem

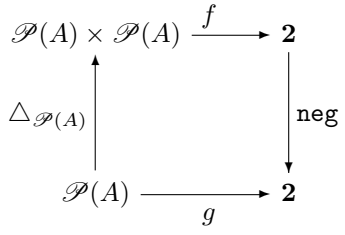
In 1997, George Boolos published another proof [1] for Cantor's Theorem, by showing that there cannot be any injection from the powerset of a set to the set. This proof has been (implicitly or explicitly) mentioned also in [7,12]. This proof is essentially Cantor's Diagonal Argument.

Theorem 2. *No function $h : \mathscr{P}(A) \to A$ can be injective.*

Proof. Let $h : \mathscr{P}(A) \to A$ be a function. Define $f : \mathscr{P}(A) \times \mathscr{P}(A) \to \mathbf{2}$ by

$$f(X, Y) = \begin{cases} 1 & \text{if } h(X) \notin Y \\ 0 & \text{if } h(X) \in Y \end{cases}$$

and let $g : \mathscr{P}(A) \to \mathbf{2}$ be the following function



Let $\mathcal{D}_h = \{a \in A \mid \exists Y \subseteq A\ (h(Y) = a\ \&\ a \notin Y)\}$. Note that for any $X \subseteq A$ we have $h(X) \notin X \longrightarrow h(X) \in \mathcal{D}_h$. We show that if h is one–to–one then g is representable by f at $\mathcal{D}_h$. For, if h is injective then for any $X \subseteq A$,

$$\begin{aligned} h(X) \in \mathcal{D}_h &\longrightarrow \exists Y \subseteq A\ (h(Y) = h(X)\ \&\ h(X) \notin Y) \\ &\longrightarrow \exists Y\ (Y = X\ \&\ h(X) \notin Y) \\ &\longrightarrow h(X) \notin X \end{aligned}$$

Whence, $h(X) \notin X \longleftrightarrow h(X) \in \mathcal{D}_h$ for all $X \subseteq A$. So, for any $X \subseteq A$,

$$\begin{aligned} f(X, \mathcal{D}_h) = 0 &\longleftrightarrow h(X) \in \mathcal{D}_h \\ &\longleftrightarrow h(X) \notin X \\ &\longleftrightarrow f(X, X) = 1 \\ &\longleftrightarrow g(X) = \mathtt{neg}(f(X, X)) = 0 \end{aligned}$$

Thus, $g(X) = f(X, \mathcal{D}_h)$. The contradiction (that `neg` possesses a fixed–point) follows as before, implying that the function h cannot be injective. ⊠

In fact the proof of the above theorem gives some more information than mere non–injectivity of any function $h : \mathscr{P}(A) \to A$, i.e., the existence of some $C, D \subseteq A$ such that $h(C) = h(D)$ and $C \neq D$.

Corollary 1. *For any function $h : \mathscr{P}(A) \to A$ there are some $C, D \subseteq A$ such that $h(C) = h(D) \in D \setminus C$ (and so $C \neq D$).*

Proof. For any $X \subseteq A$ we had $h(X) \notin X \longrightarrow h(X) \in \mathcal{D}_h$, whence $h(\mathcal{D}_h) \notin \mathcal{D}_h \longrightarrow h(\mathcal{D}_h) \in \mathcal{D}_h$, and so $h(\mathcal{D}_h) \in \mathcal{D}_h$. Thus, there exists some $\mathcal{C}_h$ such that $h(\mathcal{C}_h) = h(\mathcal{D}_h)$ and $h(\mathcal{D}_h) \notin \mathcal{C}_h$. So, for these $\mathcal{C}_h, \mathcal{D}_h \subseteq A$ we have $h(\mathcal{C}_h) = h(\mathcal{D}_h) \in \mathcal{D}_h \setminus \mathcal{C}_h$. ⊠

3 Yablo's Paradox

To counter a general belief that all the paradoxes stem from a kind of circularity (or involve some self–reference, or use a diagonal argument) Stephen Yablo designed a paradox in 1985 that seemingly avoided self–reference ([14,15]). Let us fix our reading of Yablo's Paradox. Consider the sequence of sentences $\{\mathcal{Y}_n\}_{n \in \mathbb{N}}$ such that for each $n \in \mathbb{N}$:

$$\mathcal{Y}_n \text{ is true } \Longleftrightarrow \forall k > n\ (\mathcal{Y}_k \text{ is untrue}).$$

The paradox follows from the following deductions. For each $n \in \mathbb{N}$,

$$\begin{aligned} \mathcal{Y}_n \text{ is true} &\Longrightarrow \forall k > n\ (\mathcal{Y}_k \text{ is untrue}) \\ &\Longrightarrow (\mathcal{Y}_{n+1} \text{ is untrue}) \text{ and } \forall k > n+1\ (\mathcal{Y}_k \text{ is untrue}) \\ &\Longrightarrow (\mathcal{Y}_{n+1} \text{ is untrue}) \text{ and } (\mathcal{Y}_{n+1} \text{ is true}), \end{aligned}$$

thus $\mathcal{Y}_n$ is not true. So,

$$\forall k\ (\mathcal{Y}_k \text{ is untrue}),$$

and in particular

$$\forall k > 0\ (\mathcal{Y}_k \text{ is untrue}),$$

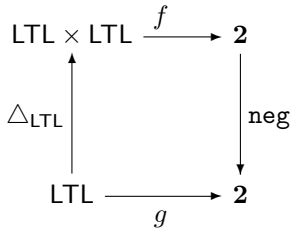
and so $\mathcal{Y}_0$ must be true (and untrue at the same time); contradiction!

3.1 Propositional Linear Temporal Logic

The propositional linear temporal logic (LTL) is a logical formalism that can refer to time; in LTL one can encode formulae about the future, e.g., a condition will eventually be true, a condition will be true until another fact becomes true, etc. LTL was first proposed for the formal verification of computer programs in 1977 by Amir Pnueli [10]. For a modern introduction to LTL and its syntax and semantics see e.g. [8]. Two modality operators in LTL that we will use are the "next" modality $\odot$ and the "always" modality $\boxdot$. The formula $\odot\phi$ holds (in the current moment) when ϕ is true in the "next step", and the formula $\boxdot\phi$ is true (in the current moment) when ϕ is true "now and forever" ("always in the future"). In the other words, $\boxdot$ is the reflexive and transitive closure of $\odot$. It can be seen that the formula $\odot\neg\phi \longleftrightarrow \neg \odot \phi$ is always true (is a law of LTL, see T1 on page 27 of [8]), since ϕ is untrue in the next step if and only if it is not the case that "ϕ is true in the next step". Also the formula $\odot\boxdot\psi$ is true when ψ is true from the next step onward, that is ψ holds in the next step, and the step after that, and the step after that, etc. The same holds for $\boxdot\odot\psi$; indeed the formula $\odot\boxdot\psi \longleftrightarrow \boxdot\odot\psi$ is a law of LTL (T12 on page 28 of [8]). Whence, we have the equivalences $\odot \boxdot \neg\phi \longleftrightarrow \boxdot \odot \neg\phi \longleftrightarrow \boxdot\neg \odot \phi$ in LTL.
Now we show the non–existence of a formula $\mathscr{Y}$ that satisfies the equivalences

$$\mathscr{Y} \longleftrightarrow \odot\boxdot\neg\mathscr{Y} \quad \left(\longleftrightarrow \boxdot\odot\neg\mathscr{Y} \longleftrightarrow \boxdot\neg \odot \mathscr{Y}\right);$$

in other words $\mathscr{Y}$ is a fixed–point of the operator $x \mapsto \odot \boxdot \neg x$ ($\equiv \boxdot \odot \neg x \equiv \boxdot\neg \odot x$); let us note that $\equiv$ stands for logical equivalence. Following [16] we can demonstrate this by the following diagram



where LTL is the set of sentences in the language of LTL and f is defined by

$$f(X,Y) = \begin{cases} 1 & \text{if } X \not\equiv \odot\boxdot\neg Y, \\ 0 & \text{if } X \equiv \odot\boxdot\neg Y. \end{cases}$$

Here, g is the characteristic function of all the Yablo–like sentences, the sentences which claim that all they say in the future (from the next step onward) is untrue.

Theorem 3. *For any arbitrary formula ϕ the formula $\left(\phi \leftrightarrow \odot \boxdot \neg\phi\right)$ is not provable in* LTL.

Proof. Assume, for the sake of contradiction, that LTL proves $\psi \leftrightarrow \odot \boxdot \neg\psi$ for some (propositional) formula ψ. For a model $\langle \mathbb{N}, \Vdash \rangle$ we consider two cases:

(i) If $m \Vdash \psi$ for some m, then $m \Vdash \odot\boxdot\neg\psi$ so $(m+1) \Vdash \boxdot\neg\psi$, hence $(m+i) \Vdash \neg\psi$ for all $i \geq 1$. In particular, $(m+1) \Vdash \neg\psi$ and $(m+j) \Vdash \neg\psi$ for all $j \geq 2$ which implies $(m+2) \Vdash \boxdot\neg\psi$ or $(m+1) \Vdash \odot\boxdot\neg\psi$ so $(m+1) \Vdash \psi$, a contradiction!

(ii) So, $k \Vdash \neg\psi$ for all k, and then $k \Vdash \odot\neg\boxdot\neg\psi$ thus $(k+1) \Vdash \neg\boxdot\neg\psi$; hence $(k+n) \Vdash \varphi$ for some $n \geq 1$, again a contradiction (with (i) above)!

So, LTL$\nvdash \big(\phi \leftrightarrow \odot\boxdot\neg\phi\big)$ for all formulas ϕ. ⊠

The above proof is very similar to Yablo's argument (in his paradox) presented at the beginning of this section, and this goes to say that Yablo's paradox has turned into a genuine mathematico–logical theorem (in LTL) for the first time in Theorem 3.

4 Priest's Inclosure Schema

In 1997 Priest [11] introduced his Inclosure Schema and showed that Yablo's paradox is amenable in it (see also [2]). In [11] Priest also shows the existence of a formula $Y(x)$ which satisfies $Y(n) \leftrightarrow \forall k > n\ \mathcal{T}(\ulcorner Y(\underline{k})\urcorner)$ for every $n \in \mathbb{N}$, where $\mathcal{T}(x)$ is a (supposedly truth) predicate; here $\ulcorner\psi\urcorner$ is the (Gödel) code of the formula ψ and for a $k \in \mathbb{N}$, $\underline{k}$ is a term representing k (e.g. $1+\cdots+1$ [$k-$times]). Rigorous proofs for the existence of such a formula $Y(x)$ (and its construction) can be found in [3,4]. Here we construct a formula $Y(x)$ which, for every $n \in \mathbb{N}$, satisfies the formula $Y(n) \leftrightarrow \forall k > n\ \Psi(\ulcorner Y(\underline{k})\urcorner)$ for some Π_1 formula Ψ, by using the Recursion Theorem (of Kleene); for recursion–theoretic definitions and theorems see e.g. [5]. Let $\mathbf{T}$ denote Kleene's T Predicate, and for a fixed Π_1 formula $\Psi(x)$ let r be the recursive function defined by $r(x,y) = \mu z\big(z > x\ \&\ \neg\Psi(\ulcorner\neg\exists u\mathbf{T}(y,\underline{z},u)\urcorner)\big)$; note that $\neg\Psi$ is a Σ_1 formula. By the S–m–n theorem there exists a primitive recursive function s such that $\phi_{s(y)}(x) = r(x,y)$; here ϕ_n denotes the unary recursive function with (Gödel) code n, so $\phi_0, \phi_1, \phi_2, \cdots$ lists all the unary recursive functions. By Kleene's Recursion Theorem, there exists some (Gödel code) e such that $\phi_e = \phi_{s(e)}$. Whence, $\phi_e(x) = \phi_{s(e)}(x) = r(x,e) = \mu z\big(z > x\ \&\ \neg\Psi(\ulcorner\neg\exists u\mathbf{T}(e,\underline{z},u)\urcorner)\big)$. So, for any $x \in \mathbb{N}$ we have $\exists u\mathbf{T}(e,\underline{x},u) \Leftrightarrow \phi_e(x)\!\downarrow \Leftrightarrow \exists z\big(z > x\ \&\ \neg\Psi(\ulcorner\neg\exists u\mathbf{T}(e,\underline{z},u)\urcorner)\big)$, or in the other words we have the equivalence $\neg\exists u\mathbf{T}(e,\underline{x},u) \leftrightarrow \forall z > x\ \Psi(\ulcorner\neg\exists u\mathbf{T}(e,\underline{z},u)\urcorner)$. Thus if we let $\mathcal{Y}(v) = \neg\exists z\mathbf{T}(e,\underline{v},z)$, then for any $n \in \mathbb{N}$ we have $\mathcal{Y}(n) \leftrightarrow \forall k > n\ \Psi(\ulcorner\mathcal{Y}(\underline{k})\urcorner)$. Let us note that Yablo's paradox occurs when Ψ is taken to be an untruth (or non–satisfaction) predicate; in fact one might be tempted to take $\neg Sat_{\Pi,1}(x,\emptyset)$ (see Theorem 1.75 of [6]) as $\Psi(x)$; but $Sat_{\Pi,1}(x,\emptyset)$ is Π_1 and so $\neg Sat_{\Pi,1}(x,\emptyset)$ is Σ_1, and our proof works for $\Psi \in \Pi_1$ only (otherwise the function r could not be recursive). Actually, the above construction shows that $Sat_{\Pi,1}(x,\emptyset)$ (in [6]) cannot be Σ_1, which is equivalent to saying that the set of true Π_1 sentences cannot be recursively enumerable, and this is a consequence of Gödel's first incompleteness theorem (cf. [3,4]).

In the following, we show that Priest's Inclosure Schema can fit in Yanofsky's framework [16]. With some inessential modification for better reading, Priest's inclosure schema is defined to be a triple $\langle\Omega,\Theta,\delta\rangle$ where

- Ω is a set of objects;
- $\Theta \subseteq \mathscr{P}(\Omega)$ is a property of subsets of Ω such that $\Omega \in \Theta$;
- $\delta : \Theta \to \Omega$ is a function such that for each $X \in \Theta$, $\delta(X) \notin X$.

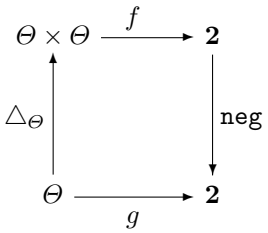
That any inclosure schema is contradictory can be seen from the fact that by the second item $\delta(\Omega)$ must be defined and belong to Ω, but at the same time by the third item $\delta(\Omega) \notin \Omega$. We show how this can be proved by the non–existence of a fixed–point for the negation function.

Theorem 4. *If an inclosure schema exists, then the negation function has a fixed–point.*

Proof. Assume $\langle \Omega, \Theta, \delta \rangle$ is a (hypothetical) inclosure schema. Define $f : \Theta \times \Theta \to \mathbf{2}$ by

$$f(X,Y) = \begin{cases} 1 & \text{if } \delta(X) \in Y \\ 0 & \text{if } \delta(X) \notin Y \end{cases}$$

Let $g : \Theta \to \mathbf{2}$ be defined as



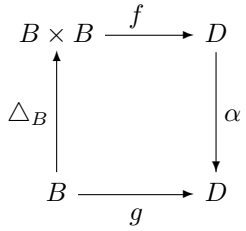
We show that g is representable by f at Ω. By the definition of δ for every $X \in \Theta$ we have $f(X, \Omega) = 1$. On the other hand by the property of δ, for any $X \in \Theta$, $\delta(X) \notin X$, and so $f(X, X) = 0$, thus $g(X) = \mathtt{neg}(f(X, X)) = 1$. Whence $g(X) = f(X, \Omega)$ for all $X \in \Theta$. Since

$$\mathtt{neg}(f(\Omega, \Omega)) = g(\Omega) = f(\Omega, \Omega),$$

so, neg has the fixed point $f(\Omega, \Omega)$ and this is a contradiction! ⊠

5 Conclusions

There are many interesting questions and suggestions for further research at the end of [16] which motivated the research presented in this paper; most of the questions remain unanswered as of today. The proposed schema, i.e., the diagram of the proof of Theorem 1,



can be used as a criterion for testing whether an argument is diagonal or not. What makes this argument (of the non–existence of a fixed–point for $\alpha : D \to D$) diagonal is the diagonal function $\triangle_B : B \to B \times B$. In most of our arguments we had $D = \mathbf{2} = \{0, 1\}$ and $\alpha = \mathtt{neg}$ by which the proof was constructed by diagonalizing out of the function $f : B \times B \to D$.

In this paper, we fit more theorems like some new proofs of Boolos for Cantor's theorem on the non-equinumerosity of a set with its powerset, a formalization of Yablo's paradox, and Priest's inclosure schema in Yanofsky's universal diagonal/fixed-point framework. For other exciting questions and examples of theorems or paradoxes which seem to be self–referential we refer the reader to the last section of [16]. It will be nice to see some of those proposals or other more phenomena fit in the above universal diagonal schema.

References

1. Boolos, G.: Constructing Cantorian Counterexamples. Journal of Philosophical Logic, 26:3, 237–239 (1997).
2. Bueno, O., Colyvan, M.: Paradox without Satisfaction, Analysis, 63:2, 152–156 (2003).
3. Cieśliński, C.: Yablo Sequences in Truth Theories, in: Kamal Lodaya (ed.), Logic and Its Applications, Proceedings of the 5th Indian Conference, ICLA 2013, Chennai, India, January 10–12, 2013, LNCS 7750, Springer, 127–138 (2013).
4. Cieśliński, C., Urbaniak, R.: Gödelizing the Yablo Sequence, Journal of Philosophical Logic, 42:5, 679–695 (2013).
5. Epstein, R.L., Carnielli, W.A.: Computability: computable functions, logic, and the foundations of mathematics, Advanced Reasoning Forum (3rd ed. 2008).
6. Hájek, H., Pudlák, P.: Metamathematics of First-Order Arithmetic, Springer (2nd. print. 1998).
7. Kanamori, A., Pincus, D.: Does GCH Imply AC Locally?, in: G. Halasz & L. Lovasz & M. Simonovits & V.T. Sós (eds.) Paul Erdös and His Mathematics II, Bolyai Society for Mathematical Studies, Vol. 11, Springer, 413–426 (2002).
8. Kröger, F., Merz, S.: Temporal Logic and State Systems, Springer (2008).
9. Lawvere, F.W.: Diagonal arguments and cartesian closed categories, Category theory, homology theory and their applications II, (Seattle Research Center of the Battelle Memorial Institute), LNM 92, Springer, 134–145 (1969).
10. Pnueli, A.: The Temporal Logic of Programs, in: Proc. 18th Ann. Symp. Foundations of Computer Science (SFCS'77) IEEE Computer Society, Washington DC, USA, 46–57 (1977).
11. Priest, G.: Yablo's Paradox, Analysis, 57:4, 236–242 (1997).
12. Raja, N.: Yet Another Proof of Cantor's Theorem, in: J.-Y. Béziau & Al. Costa-Leite (eds.) Dimensions of Logical Concepts, Coleção CLE: Volume 54, 209–217 (2009).
13. Russell, B.: On some difficulties in the theory of transfinite numbers and order types, Proceedings of the London Mathematical Society, vol. s2–4, no. 1, 29–53 (1907).
14. Yablo, S.: Truth and Reflection, Journal of Philosophical Logic, 14:3, 297–349 (1985).
15. Yablo, S.: Paradox without Self-Reference, Analysis, 53:4, 251–252 (1993).
16. Yanofsky, N.S.: A Universal Approach to Self-Referential Paradoxes, Incompleteness and Fixed Points, Bulletin of Symbolic Logic, 9:3, 362–386 (2003).

Quasi-Bayesian Belief Revision on Spohn Plausibility Structures

Ciyang Qing

ILLC, University of Amsterdam

Abstract Previous works [8,6,7,3] established the connection between belief revision and formal learning theory and it has been shown that the classical AGM belief revision framework has restricted universal learning power [3]. Inspired by works from the *Bayesian concept learning* literature [10], we propose Quasi-Bayesian belief revision as an alternative non-AGM method and prove that it has stronger universal learning power. We argue that AGM belief revision is too conservative and ignores an aspect of parsimony, i.e., one should gradually give up complex hypotheses whose surplus predictions have never been verified, in favor of simpler ones that explain all observations so far equally well.

1 Introduction

When an agent observes a new piece of information, how should it change its beliefs? If the new piece of information is consistent with the current belief, then it seems very natural to simply combine them to form a new belief. In the classical AGM belief revision framework [1], this intuition is implemented in terms of expansion. It states that if the new information is consistent with the current belief, then the new belief should be the logical closure of the conjunction of the current belief and the new information. One consequence is that everything in the current belief would remain. In particular, if the new observation is already currently believed, then the AGM procedure would leave the current belief totally unchanged.

This revision method for consistent information is so natural and intuitively plausible that it tends to be taken for granted and receives fairly little attention. Traditionally, the belief revision literature concerns situations in which the new piece of information contradicts the current belief, and develops various revision methods to deal with the complexity and subtlety in these situations.

Previous works [8,6,7,3] establish the connection between belief revision and formal learning theory and show the limitation of AGM belief revision methods in terms of universal learning power. In light of these results, in this paper we re-examine the case of new consistent information and argue that the ostensibly impeccable AGM revision postulate is too conservative and lacks an important aspect of parsimony, i.e., if a simpler set of beliefs could explain everything that has been observed equally well, then one should probably discard more complex beliefs in favor of it.

We will formalize this idea and provide a non-AGM belief revision policy inspired by works from the *Bayesian concept learning* literature [10]. We will prove that it has better universal learning power, in the sense of truth-tracking in the limit [3].

The rest of the paper is organized as follows. In Section 2 we introduce the semantic approach to belief and belief revision, and recent work on using formal learning theory to evaluate belief revision methods [3]. In Section 3 we review works on Bayesian concept learning [10], with an emphasis on the role *likelihoods* play in belief revision and learning. In Section 4 we introduce Quasi-Bayesian revision that incorporates likelihoods in a qualitative style on Spohn Plausibility Structures [9]. We then introduce the notion of (ω-)strong universality and prove that Quasi-Bayesian revision is ω-strongly universal in Section 5.

2 Belief Revision and Truth-Tracking Universality

In this section, we describe a model-theoretic approach to the semantics of belief and formalize the problem of belief revision. Then we introduce the idea of evaluating belief revision policies in light of formal learning theory [5], following the setting in recent literature [3].

Definition 1. *An* epistemic space (S, Φ) *consists of a countable set* S *of* epistemic states *(also called possible worlds), and a family of* observable properties $\Phi \subseteq \mathcal{P}(S)$. *A* plausibility space $(S, \Phi, \leq)$ *is an epistemic space* (S, Φ) *paired with a* plausibility order $\leq$, *which is a preorder on* S. *For two states* $w, s \in S$, $w \leq s$ *means* w *is more plausible*[1] *than* s.

The semantics of belief is defined as follows: in epistemic state s, *a property (also called a proposition)* $\varphi \in \mathcal{P}(S)$ *is believed in* s *iff there exists some* $w \leq s$ *such that for any* $v \leq w$, $v \in \varphi$. *Formally, we write* $s \vDash B\varphi$.

Note that the property φ here need not be observable. Also, in many situations the plausibility order $\leq$ is well-founded, in which case it can be shown that we have an equivalent definition: $s \vDash B\varphi$ iff we have $w \in \varphi$ for every most plausible world $w \in \min_{\leq} S$.

Example 1. Consider an epistemic space (S, Φ), with $S = \mathbb{N}, \Phi = \{\{0, \ldots, k\} \mid k \in \mathbb{N}\}$, as depicted below:

$$[\ldots[[[0\ ]\ 1\ ]\ 2\ ]\ldots]\ .$$

If we take the plausibility order $\leq_{\mathbb{N}}$ to be the order of natural numbers, i.e., $0 <_{\mathbb{N}} 1 <_{\mathbb{N}} 2 < \cdots$, then 0 is the most plausible world and we have for instance $0 \vDash B\{0,1,2\}$, $0 \vDash B\{0,3\}$, $2 \vDash B\{0,3\}$, and $2 \nvDash B\{1,2\}$.

[1] At first sight it might seem counter-intuitive that $w \leq s$ means w is more plausible. The main reason is that we often want to have a most plausible world, but traditionally well-foundedness is defined in terms of the existence of a least element w.r.t. $\leq$. An alternative formulation, which is very common in the literature, is to call $\leq$ the implausibility order and then $w \leq s$ means, literally, that w is less implausible than s, which just means w is more plausible than s.

These examples are meant to illustrate and emphasize that (1) in this semantics one can talk about unobservable properties (e.g., $\{0, 3\}$) as beliefs, and (2) the truth of a belief statement does not depend on where it is evaluated, but rather the overall structure of the plausibility space. In particular it depends on the set of most plausible worlds (for well-founded plausibility orders). Hence one can study a belief set (more precisely, a belief closure) in terms of the underlying plausibility structure.

The reason why we distinguish observable properties from properties in general is to capture the intuition that not all properties need to be directly observable and a belief revision policy may be defined to handle only those properties that are directly observable. From the above definition of belief, we can see that the set of beliefs is determined by the plausibility space. Thus, in the model-theoretic approach, a belief revision policy π is taken to be a family of transformations on plausibility spaces.

Definition 2. *A belief revision policy π is a family of transformations such that for any plausibility space $(S, \Phi, \leq)$, any observable property $\sigma \in \Phi$, $\pi(\sigma)$ maps the plausibility space $(S, \Phi, \leq)$ to a new plausibility space $(S^\sigma, \Phi^\sigma, \leq^\sigma)$, where $S^\sigma \subseteq S$ is a subset of the original epistemic states, $\Phi^\sigma = \{P \cap S^\sigma \mid P \in \Phi\}$ is the set of observable properties restricted in the new set of epistemic states S^σ, and $\leq^\sigma$ is a new plausibility order defined on S^σ.*

Hence, in order to define a belief revision policy, we need to specify two things for each plausibility space $(S, \Phi, \leq)$ and $\sigma \in \Phi$: first, the new set of epistemic states S^σ (and Φ^σ is thus determined), and second, the new plausibility order $\leq^\sigma$ on S^σ.

Example 2. Conditioning is a belief revision policy such that for any plausibility space $(S, \Phi, \leq)$ and any observable property $\sigma \in \Phi$, we let $S^\sigma = S \cap \sigma = \sigma$ and $\leq^\sigma$ be the restriction of $\leq$ on S^σ. In other words, conditioning throws away all epistemic states that are inconsistent with σ, keeping the plausibility relation among the remaining states the same. *Probabilistic conditioning* is a special case of conditioning, which is only defined for plausibility spaces where there exists a probability distribution $p(S)$ on S (for the purposes of this paper, it suffices to think of it as a function that maps every state s to a non-negative number $p(s)$ such that $\sum_{s \in S} p(s) = 1$) such that for any $w, v \in S$, $w \leq v$ iff $p(w) \geq_{\mathbb{R}} p(v)$ ($\geq_{\mathbb{R}}$ is the normal order for numbers). Probabilistic conditioning is important in probability theory, but for the present paper the only relevant thing is that a plausibility order derived from a probability distribution is always well-founded.

Many belief revision policies have been defined in the literature, and a natural question is how we should evaluate these policies. One perspective is that they can be evaluated in terms of their ability to track the truth in the limit. Previous works [8,6,7,3] have established such a connection between belief revision and formal learning theory. The general idea is that suppose there is a real world $s \in S$ and the nature keeps generating informative observations about s, then

an ideal belief revision policy should allow us to keep revising our belief using the observations and ultimately enable us to identify the real world. Thus two belief revision methods can be compared in terms of whether or not they can meet this requirement, or more precisely, under how strong additional assumptions they can meet this requirement. In the sequel we will introduce the formalization of this intuition and review some previous results [3].

First we formally define what it means to learn the truth from observations.

Definition 3. *Given an epistemic space (S, Φ), the observable*[2] property set *of a state $s \in S$, denoted as Φ_s, is defined as $\{P \in \Phi \mid s \in P\}$, i.e. the set of all its properties. We say a* data stream *$\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots)$, which is an infinite sequence of properties, as* sound *and* complete *w.r.t. a state s iff it totally comes from and exhausts Φ_s, i.e. $\{\varepsilon_n \mid n \in \mathbb{N}\} = \Phi_s$.*

Definition 4. *A learning method L for (S, Φ) is a function that takes in a finite sequence of observable properties and outputs a* hypothesis *which is a subset of S. A state s is learnable by L iff for every sound and complete data stream ε w.r.t. s, there exists an N s.t. $L(\varepsilon_1, \ldots, \varepsilon_n) = \{s\}$ for all $n \geq N$. An epistemic space (S, Φ) is* learnable by *L iff every state in S is learnable and an epistemic space* learnable *iff there exists a learning method L by which the epistemic space is learnable.*

The above definitions formalize the process of learning from informative observations to track down the truth. Obviously not every epistemic state is learnable. For instance, if two states $s, s' \in S$ have the same property set, i.e., $\Phi_s = \Phi_{s'}$, then there is no way that the learner can distinguish between s and s' from observations. Since we want to evaluate belief revision methods in terms of their learning power, we will only consider those epistemic spaces that are learnable (by any learner at all). Note that this implies that for these spaces if $s \neq s'$, then $\Phi_s \neq \Phi_{s'}$ (but note that $\Phi_s \neq \Phi_{s'}$ for $s \neq s'$ does not guarantee learnability).

Next we illustrate how to use a belief revision policy π to derive a learning method. Given an epistemic space (S, Φ) and a *prior* plausibility order $\leq_0$ on S, we have a plausibility space $(S, \Phi, \leq_0)$. Now we can derive a learning method $L_\pi^{\leq_0}$ from belief revision policy π with prior $\leq_0$ as follows. For a finite sequence of observables $(\varepsilon_1, \ldots, \varepsilon_n)$, we first use the belief revision policy on ε_1 to transform $(S, \Phi, \leq_0)$ into a new plausibility space $(S^{\varepsilon_1}, \Phi^{\varepsilon_1}, \leq_0^{\varepsilon_1})$, which we simply denote as $(S_1, \Phi_1, \leq_1)$. Then we use the belief revision policy again (on ε_2) to transform $(S_1, \Phi_1, \leq_1)$ into $(S_1^{\varepsilon_2}, \Phi_1^{\varepsilon_2}, \leq_1^{\varepsilon_2})$, denoted as $(S_2, \Phi_2, \leq_2)$, and so on, until we have used the belief revision policy on ε_n and obtain $(S_n, \Phi_n, \leq_n)$. Now the output of the learning method $L_\pi^{\leq_0}$ derived from π with prior $\leq_0$ is defined as $L_\pi^{\leq_0}((\varepsilon_1, \ldots, \varepsilon_n)) = \min_{\leq_n} S_n$. (Note that if $\leq_n$ is not well-founded then $L_\pi^{\leq_0}$ outputs the empty set for this finite sequence, but this does not violate our definition.) Intuitively, once a prior plausibility order $\leq_0$ is given, we can use the belief revision policy iteratively on the finite data sequence and output the set of best worlds in the final plausibility space as the hypothesis.

[2] In the rest of the paper, all properties are by default assumed to be observable, unless explicitly stated otherwise.

Hence we can evaluate belief revision policies in terms of their learning powers.

Definition 5. *A belief revision policy π is* universal[3] *iff for every learnable epistemic space, there exists a prior $\leq_0$ such that $L_\pi^{\leq_0}$ can learn that space. The policy is* standardly universal *iff it is universal and the prior $\leq_0$ is required to be well-founded. A policy is* strongly universal *iff $L_\pi^{\leq_0}$ can learn the space for all priors.*

It is not hard to see that strong universality implies standard universality, which in turn implies universality. Within the AGM framework, [3] proved that while there do exist universal belief revision mechanisms (e.g., conditioning), there is no standardly universal belief revision method (which means, for instance, probabilistic conditioning is not universal). However, it seems that universality is a rather weak, since the success of learning crucially depends on the prior one has, and normally our prior belief is constructed independent of the belief revision policy so one cannot guarantee that they will fit together.

3 Bayesian Concept Learning

In the previous section we see that probabilistic conditioning is not universal. This result is sometimes phrased as Bayesian conditioning is not universal. While this is only a matter of terminology, typically the term "Bayesian" implies the use of a *likelihood* function together with prior, rather than simple probabilistic conditioning. In this section we will briefly review the Bayesian framework for concept learning [10] and discuss several implications. We will propose a new belief revision method inspired by this framework in the next section.

Bayesian concept learning is closely related to the *set learning* paradigm in the formal learning theory literature, but with slightly different focus (hence in the sequel we will reformulate the details to fit the theme of this paper). The central question is that, given a class of languages $\mathcal{L} = (L_i)_{i\in\mathbb{N}}$ and an instance s from some target language $L_t \in \mathcal{L}$, what can we learn about this target language? Using the Bayes rule we have

$$p(L_i \mid s) = \frac{\Pr(L_i) \cdot p(s \mid L_i)}{\sum_{j\in\mathbb{N}} \Pr(L_j) \cdot p(s \mid L_j)}, \tag{1}$$

where $p(L_i \mid s)$ is the *posterior* probability that L_i is the target language given the sample s, $\Pr(L_i)$ is the *prior* probability that L_i is the target language, and $p(s \mid L_i)$ is the *likelihood* that we will observe s if L_i is the target language. It is the likelihood term that plays a key role in learning in this framework. In the simpliest cases, in which all the languages are finite, one can use the *strong sampling* assumption which postulates that the sample instance s is drawn

[3] The definitions of universality and standard universality follow from [3].

uniformly from the extension of the target language, thus if $s \in L_i$ the likelihood term $p(s \mid L_i)$ is the inverse of the size of the L_i:

$$p(s \mid L_i) = \frac{1}{|L_i|} \text{ if } s \in L_i, \text{ otherwise } 0 \ . \tag{2}$$

This equation is also referred to as the *size principle*, as it favors languages (consistent with the sample s) that have smaller sizes. As evidence accumulates, i.e. after observing several samples and updating the posterior probabilities accordingly, the differences in the likelihood terms might override the prior orderings derived from the prior probabilities.

Example 3 (adapted from [10]). Suppose the class of languages consists of all the (non-empty) subsets of numbers from 1 to 100. After observing 2 and 4, we might guess the target language is the one that has all even numbers, but after observing 2, 4, 8, 16 and 32 we will probably come to believe the target language is the powers of 2, even though the additional observations (8, 16, and 32) are all even numbers, which are perfectly consistent with the original hypothesis.

Let us examine this example in detail within the belief revision framework introduced in the previous section and the Bayesian concept learning framework in this section, to illustrate the potential limitation of the conservative AGM revision policies.

An epistemic state s in this case is a non-empty subset of numbers from 1 to 100. For instance, we have an epistemic state $s_o = \{1, 3, \ldots, 99\}$ consisting of all the odd numbers. Similarly, we have an epistemic state $s_e = \{2, 4, \ldots, 100\}$ consisting of all the even numbers and another epistemic state $s_p = \{2, 4, 8, 16, 32, 64\}$ consisting of all powers of 2.

The first observation of 2 is technically the set of all epistemic states that contain 2, i.e., $\varepsilon_1 = \{s \mid 2 \in s\}$. In particular we have $s_o \notin \varepsilon_1$ and $s_e, s_p \in \varepsilon_1$ Simililarly we have $\varepsilon_2 = \{s \mid 4 \in s\}$, $\varepsilon_3 = \{s \mid 8 \in s\}$, and so on.

After observations of 2 and 4, one comes to believe that the target language is s_e, which means s_e is the least element in the plausibility order. Now it can be checked that $s_e \vDash B\varepsilon_3$, i.e., the agent already believes that 8 is in the target language. According to the AGM postulates, the observation of 8 will not change the belief, nor would any subsequent observations of 16, 32, and so on. Thus the agent has to hold on to the belief that the target language is the set of even numbers, even if he only observes powers of 2 for thousands of times. Clearly this result is counter-intuitive. Moreover, it is doubtful whether such a conservative revision policy is rational, since the agent ends up with lots of false belief, e.g., that 14 is in the target language.

The main problem with the conservative AGM belief revision policy in this example is that it misses an important aspect of parsimony, i.e., one should give up unnecessary beliefs if observations can be equally explained with or without them. In our example, the additional belief that 14 is in the target language does not help explain the observations and thus is unnecessary. In addition, the fact

that 14 has never been observed so far provides reasons to doubt and even drop the belief that 14 is in the target language.

On the other hand, the Bayesian framework can overcome the above problem by making use of the likelihood term and size principle. Even though the observation of 8 is consistent with the current belief s_e, the likelihood term $p(8 \mid s_e) = \frac{1}{|s_e|}$ is only 1/50. The alternative competing hypothesis s_p has a much higher likelihood term $p(8 \mid s_p) = \frac{1}{|s_p|} = 1/6$ because it predicts much fewer numbers in the target language. This difference in likelihood means that as observations accumulate, the more specific alternative can override the current belief, even if these observations do not directly contradict the current belief.

Admittedly, there is always a risk that the revised belief turns out to be wrong. For instance, in the previous example the target language could indeed be the set of even numbers and 14 could turn out to be the sixth observation which refutes the revised belief that only powers of 2 are in the target language. However, in the long run such a risk is under control and worth taking, because we know that if we are wrong then at some later stage our observation will reveal the mistake and we can further revise our belief. In contrast, if we are too conservative to take any risk, we may get stuck and lose the possibility of learning the truth.

The non-conservativity in Example 3 has been shown robust in the context of human concept learning and can be captured by the above Bayesian framework. The implications of this line of research are the following: First and foremost, non-conservativity is not scarce, nor is it necessarily irrational, so the conservatism prescribed by the AGM framework [1] should be carefully evaluated rather than taken for granted. Secondly, such non-conservativity may be helpful to long-term learning, as it can help us override false priors. Finally, the likelihood is crucial to overriding the prior and thus it is worth trying to incorporate it in a belief revision method.

However, if we want to formally evaluate the idea of using likelihoods in belief revision in terms of its learning power, the Bayesian framework needs to be adapted to fit a qualitative setting. For instance, the strong sampling assumption underlying the size principle is not satisfied as the data are only required to be complete. For this purpose, we will use Spohn Plausibility Structures [9] to carry out a qualitative version of Bayesian revision. There are two reasons that Spohn Plausibility Structures are particularly suitable. First of all, using ordinals to measure degrees of implausibility gives us structures fine-grained enough to gradually integrate different sources of information. Secondly, using ordinals automatically ensures that the plausibility ordering is always well-founded.

4 Spohn Plausibility Structures and Quasi-Bayesian Revision

As noted in the previous section, Spohn Plausibility Structures are extensions of plausibility structures introduced in Definition 1, where each state is associated with a *degree of implausibility*.

Definition 6. *A* Spohn plausibility structure (S, Φ, g) *is an epistemic space together with an implausibility function* g *that assigns every state* $s \in S$ *to some ordinal* $g(s)$ *which intuitively means the degree of implausibility of* s*. Note that we can easily derive a corresponding plausibility space* $(S, \Phi, \leq)$ *from* (S, Φ, g)*, i.e. let* $s \leq t$ *iff* $g(s) \leq g(t)$*.*

Since g maps states to ordinals, the derived plausibility relation is always well- founded, and thus belief state(s) can be simply defined as state(s) with the least degree of implausibility.

We can similarly adapt the definition of a belief revision policy.

Definition 7. *A belief revision policy for Spohn structures is a map from a Spohn plausibility structure* (S, Φ, g) *and a property* $\sigma \in \Phi$ *to a new Spohn plausibility structure* $(S^\sigma, \Phi^\sigma, g^\sigma)$*, where* $S^\sigma \subseteq S$ *and* $\Phi^\sigma = \{P \cap S^\sigma \mid P \in \Phi\}$ *and* g^σ *is a new implausibility function defined on* S^σ*. Hence a belief revision policy is defined by specifying* S^σ *and* g^σ *for each* $\sigma \in \Phi$*.*

Now we define the Quasi-Bayesian revision policy.

First, for a state $s \in S$, define its *limit property* $\varphi_s = \bigcap \Phi_s$. Note that by definition we always have $s \in \varphi_s$, but φ_s might not be an observable property!

Second, for a property $\sigma \in \Phi$ and a state s, the *ambiguity* of σ for s, denoted as $\alpha(\sigma \mid s)$ is defined as the cardinality of the set $\{s' \in \sigma \mid \varphi_s \subset \varphi_{s'}\}$, if it is less than K, or K if such a cardinality is greater than K (including infinite), where K is a fixed positive natural number as a free parameter of the policy[4]. The intuition is that if for some other s', which is more general (having fewer properties, note the proper subset relation in the definition) than s, s' is also consistent with the observation, then σ is ambiguous for s because it could have been more specific to rule out s'. The more such states as s', the more ambiguous σ is for s.

Finally, we can define the quasi-Bayesian revision policy as follows: for (S, Φ, g) and $\sigma \in \Phi$, let $S^\sigma = S \cap \sigma = \sigma$, which is the same as conditioning, and $g^\sigma(s) = g(s) + \alpha(\sigma \mid s)$, which is the qualitative counterpart of the size principle as states that are more ambiguous become less plausible.

To summarize, we have the following definition for quasi-Bayesian revision.

Definition 8. *For* (S, Φ, g) *and* $\sigma \in \Phi$*, the quasi-Bayesian revision policy transforms* (S, Φ, g) *into* $(S^\sigma, \Phi^\sigma, g^\sigma)$*, where* $S^\sigma = \sigma$*,* $\Phi^\sigma = \{P \cap S^\sigma \mid P \in \Phi\}$ *and* $g^\sigma(s) = g(s) + \alpha(\sigma \mid s)$*.*

Here $\alpha(\sigma \mid s) = \min\{|\{s' \in \sigma \mid \varphi_s \subset \varphi_{s'}\}|, K\}$ *and* $\varphi_s = \bigcap\{P \in \Phi \mid s \in P\}$*.*

Example 4. Let us turn back to Example 1, which is used to show that probabilistic conditioning cannot be universal in [3], to see how the quasi-Bayesian revision works and get an intuition on how it might transcend the limitation of simple probabilistic conditioning.

[4] At this point it might be unclear why the ambiguity is dealt with this way when there are infinitely such states instead of being set to ω. It is mainly for some technical reason that will become clear in the later proof.

It is easy to see that $\varphi_i = \{0, \ldots, i\}$ for $i \in \mathbb{N}$ so $\varphi_0 \subset \varphi_1 \subset \varphi_2$. Now suppose we have $g(s) = s$ and observe the property $\sigma = \{0, 1, 2\}$. Then according to the definition we have $S^\sigma = \{0, 1, 2\}$, $\Phi^\sigma = \{\{0\}, \{0, 1\}, \{0, 1, 2\}\}$, $\alpha(\sigma \mid 0) = 2, \alpha(\sigma \mid 1) = 1, \alpha(\sigma \mid 2) = 0$ and thus $g^\sigma(0) = g(0) + \alpha(\sigma \mid 0) = 2$ and likewise $g^\sigma(1) = g^\sigma(2) = 2$. Note that if we update again with σ, the epistemic space remains the same, and $g^{\sigma\sigma}(0) = 4 > g^{\sigma\sigma}(1) = 3 > g^{\sigma\sigma}(2) = 2$. Thus we see how the prior plausibility relation $0 < 1 < 2$ can be overriden even though the observation is consistent with all the states. It should also be clear why simple probabilistic conditioning would fail, i.e., it is too conservative to override the prior $0 < 1 < 2$, even when the purported difference between state 0 and the actual state 2, i.e., $\{0\}$ and $\{0, 1\}$, never appear after millions of observations.

5 Strong Universality of Quasi-Bayesian Revision

Having introduced the Quasi-Bayesian revision method, we will establish its learning power in this section.

Definition 9. *A belief revision policy π is* ω-strongly universal *iff for every learnable epistemic space and any one-to-one prior whose range is a subset of $\mathbb{N}$, the derived learning method can learn the space*[5].

Even though ω-strong universality is still weaker than strong universality, e.g., it still might not work for the unbiased prior on $\mathbb{N}$, in practice it is usually close enough to strong universality.

We will prove that Quasi-Bayesian is ω-strongly universal. To prove this, we need to first introduce the tell-tale set theorem [2] as a lemma.

Lemma 1. *If an episemic space (S, Φ) is learnable, then for each state $s \in S$, there exists a finite subset $D_s \subseteq \Phi_s$ s.t. if $D_s \subseteq \Phi_{s'} \subseteq \Phi_s$ then it must be $s' = s$. Such a D_s is called a tell-tale set for s.*

Theorem 1. *Quasi-Bayesian revision is ω-strongly univeral.*

Proof. For any learnable epistemic space (S, Φ), any prior g_0 which is a one-to-one mapping from S to $\mathbb{N}$, and any state $s \in S$ and any data-stream ε which is sound and complete w.r.t. s, from the lemma we know s has a finite tell-tale set $D_s \subseteq \Phi_s$. Since ε is sound and complete, there exists an N s.t. $D_s \subseteq \{\varepsilon_i \mid 1 \leq i \leq N\} \subseteq \Phi_s$. Thus after N quasi-Bayesian revisions we have $S_N = \bigcap_{i=1}^{N} \varepsilon_i$. Suppose there is an $s' \in S_N$ s.t. $\varphi_s \subseteq \varphi_{s'}$. By definition we have $s \in \varphi_s \subseteq \bigcap \Phi_{s'}$ which means that for every $P \in \Phi_{s'}$ we have $s \in P$ and thus $P \in \Phi_s$. Hence we have $\Phi_{s'} \subseteq \Phi_s$. On the other hand since $s' \in S_N$ we know $s' \in \varepsilon_i$ and thus $\varepsilon_i \in \Phi_{s'}$ for $1 \leq i \leq N$, which means $D_s \subseteq \{\varepsilon_i \mid 1 \leq i \leq N\} \subseteq \Phi_{s'}$. This means that $D_s \subseteq \Phi_{s'} \subseteq \Phi_s$ which according to the definition of a tell-tale set yields that $s' = s$. This means that $\alpha(\varepsilon_n \mid s) = 0$ for all $n > N$, which in turn means

[5] Again, the requirement of one-to-one is more technical than conceptual and may be relaxed, e.g., to finite-image, but we assume it to simplify the proof.

that $g_n(s) = g_N(s)$ for all $n > N$. Note that $g_N(s)$ is finite (because at each step it can increase at most K), g_0 is one-to-one, and for any t, $g_i(t)$ is increasing w.r.t. i, hence at stage N there can be only finitely many states whose degrees of implausibility are not greater than s. There are two types of such states. First, if a state $s' \notin \varphi_s$ then there exists some $P \in \Phi_s$ s.t. $s' \notin P$, and after finite amount of time P will appear in the data-stream (since it is complete) and thus s' will be deleted. Second, for a state $s' \in \varphi_s$ ($s' \neq s$), we know that $\Phi(s) \subset \Phi(s')$ (the inequality holds because $\Phi(s) = \Phi(s')$ would imply $s = s'$ for otherwise it would be impossible for the learner to distinguish s and s') and thus we have $\varphi_{s'} \subset \varphi_s$ (because there is some $P \in \Phi(s')$ which is not in $\Phi(s)$, thus $s \notin P \subseteq \varphi_{s'}$, which shows the inequality does hold.) Hence we know that $\alpha(\varepsilon_n \mid s') > 0$ for all $n \geq N$, which again means that after some finite amount of time the degree of implausibility of s' will exceed that of s. Hence after finitely many steps the true state s will be the only state among the remaining states that has the smallest degree of implausibility (which never changes since stage N) and it will stay as the most plausible world afterwards. This finishes the proof that quasi-Bayesian belief revision is $\omega-$strongly universal. □

6 Conclusion and Future Work

We argued that the AGM framework is too conservative in that it ignores an important aspect of parsimony, i.e., giving up hypotheses whose surplus predictions have never been verified, in favor of simpler ones that can explain the observations so far equally well. We illustrate that it can be problematic even when the new information is consistent with the current belief, since an agent using an AGM belief revision policy may be too conservative to learn the truth from observations.

Previous works concerning the connection between belief revision and formal learning theory [8,6,7,3] have pointed out the limitation of AGM policies in terms of learning power and the general tension between conservatism and learnability. Nevertheless, the focus has been mainly on constructing better prior plausibility orders and choosing among revision methods for new information that contradicts the current belief within the AGM framework.

We extended these previous works by considering cases of revision with consistent new information. Inspired by Bayesian concept learning, we proposed quasi-Bayesian as an alternative non-AGM belief revision method and proved that it is ω-strongly universal, a good property of learnability. This result sheds new light on the tradeoff between conservativity and learnability.

There are several possible extensions of the current work. First, though the K parameter in the policy is used in the proof, it seems rather artificial. Future work should establish whether this requirement is necessary, and if yes, what the philosophical implication would be. Second, we need to better understand the difference between ω-strong universality and strong universality in general. Some optimizations in the proof are easy, as mentioned earlier, the precise boundary is nevertheless unclear. Thirdly, the current quasi-Bayesian belief revision method

is based on Spohn Plausibility Structures, which is a specific instance of a general framework of *plausibility measures* [4]. It remains to be seen whether the quasi-Bayesian belief revision method can be generalized. Finally, [3] also takes into account cases where the data might include finitely many errors which are later on corrected and shows that certain methods can be universal for that type of data. It would be interesting to see how to generalize Quasi-Bayesian to maintain ω-strong universality.

Acknowledgements Thanks to Nina Gierasimczuk and Alexandru Baltag for advising this project, and to the ESSLLI reviewers for the helpful comments.

References

1. Alchourrón, C.E., Gärdenfors, P., Makinson, D.: On the logic of theory change: Partial meet contraction and revision functions. The Journal of Symbolic Logic 50(2), pp. 510–530 (1985), http://www.jstor.org/stable/2274239
2. Angluin, D.: Inductive inference of formal languages from positive data. Information and control 45(2), 117–135 (1980)
3. Baltag, A., Gierasimczuk, N., Smets, S.: Belief revision as a truth-tracking process. In: Proceedings of the 13th Conference on Theoretical Aspects of Rationality and Knowledge. pp. 187–190. TARK XIII, ACM, New York, NY, USA (2011), http://doi.acm.org/10.1145/2000378.2000400
4. Friedman, N., Halpern, J.Y.: Plausibility measures and default reasoning. Journal of the ACM 48(4), 648–685 (2001)
5. Gold, E.M.: Language identification in the limit. Information and control 10(5), 447–474 (1967)
6. Kelly, K.: The learning power of belief revision. In: Gilboa, I. (ed.) Proceedings of the 7th conference on Theoretical aspects of rationality and knowledge. pp. 111–124. Morgan Kaufmann Publishers Inc. (1998)
7. Kelly, K.: Iterated belief revision, reliability, and inductive amnesia. Erkenntnis 50(1), 7–53 (1999)
8. Kelly, K., Schulte, O., Hendricks, V.: Reliable belief revision. In: Luisa, M., Chiara, D. (eds.) Logic and Scientific Methods, pp. 383–398. Dordrecht Kluwer (1997)
9. Spohn, W.: Ordinal conditional functions: A dynamic theory of epistemic states. In: Harper, W.L., Skyrms, B. (eds.) Causation in Decision, Belief Change, and Statistics, The University of Western Ontario Series in Philosophy of Science, vol. 42, pp. 105–134. Springer Netherlands (1988), http://dx.doi.org/10.1007/978-94-009-2865-7_6
10. Tenenbaum, J.B.: A Bayesian framework for concept learning. Ph.D. thesis, Massachusetts Institute of Technology (1999)

Diffusion, Influence and Best-Response Dynamics in Networks: An Action Model Approach

Rasmus K. Rendsvig
rendsvig@gmail.com
LUIQ, Lund University, Sweden

Abstract Threshold models and their dynamics may be used to model the spread of 'behaviors' in social networks. Regarding such from a modal logical perspective, it is shown how standard update mechanisms may be emulated using action models – graphs encoding agents' decision rules. A small class of action models capturing the possible sets of decision rules suitable for threshold models is identified, and shown to include models characterizing best-response dynamics of both coordination and anti-coordination games played on graphs. We conclude with further aspects of the action model approach to threshold dynamics, including broader applicability and logical aspects. Hereby, new links between social network theory, game theory and dynamic 'epistemic' logic are drawn.

An individual's choice of phone, language use or convictions may be influenced by the people around her [12,22,23]. How a new trend spreads through a population depends on how agents are influenced by others, which in turn depends on the structure of the population and on how easy agents are to influence.

This paper focuses on one particular account of social influence, the notion of 'threshold influence' [15]. Threshold influence relies on a simple imitation or conformity pressure effect: agents adopt a behavior/fashion/semantics whenever some given threshold of their social network neighbors have adopted it already. So-called *threshold models*, introduced by [11,19], represent diffusion dynamics under threshold influence. Threshold models have received much attention in recent literature [10,14,16,21], also from authors in the logic community [1,6,7,15,18,20,24].

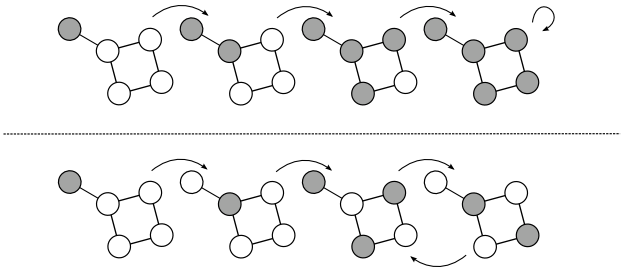


Fig. 1. (Definitions below): A threshold model with 5 agents, threshold $\theta = \frac{1}{4}$, and behavior B marked by gray. Top: agents change behavior in accordance with equation (1) and the dynamics reach a fixed point. Bottom: agents update according to equation (2). Here, the dynamics loop.

In this paper, a novel approach to threshold models is taken by constructing the dynamics using action models and product update [3,4,9]. In this context, an action model may be regarded as a graph that encodes decision rules. The product of a threshold model and an action model is again a threshold model, but where each agent has now updated their behavior according to the encoded decision rules.

The paper progresses as follows. First, threshold models and two typical update rules are introduced. We then introduce a modal language interpreted over threshold models, along with action models and product update. We produce an action model for each of the two introduced update rules, and show the stepwise equivalence of the two approaches. These two action models gives rise to a small class of action models, which is investigated in relation to tie-breaking rules, coordination game and anti-coordination game best-response dynamics. We conclude with a discussion of further aspects of the action model approach to threshold dynamics, including broader applicability and logical aspects.

The motivation for the work is primarily technical. The author found it interesting that threshold dynamics could so straightforwardly be encoded using action models. There is however an interesting conceptual twist: action models are not interpreted as being *informational events*, but as encoding *decision rules* of agents. Hence, the class arising from the action model encoding best-responses in coordination games may be seen as containing all possible sets of decision rules compatible with agents acting under the used notion of threshold influence. The class contains variations of tie-breaking rules, and shows a neat symmetry: for each "coordination game action model", the class contains a "dual" version for anti-coordination games. From a logical perspective, this class is interesting as each arising dynamics may be treated in a uniform manner, using the reduction axiom method well-known from dynamic epistemic logic [3,8].

1 Threshold Models and their Dynamics

Threshold Models. A threshold model includes a network N of agents $\mathcal{A}$ and a behavior B (or fashion, or product, or viral video) distributed over the agents. As such, it represents the current spread of B through the network. An adoption threshold prescribes how the state will evolve: agents adopt B when the proportion of their neighbors who have already adopted it meets the threshold. Formally, a threshold model is a tuple $\mathcal{M} = (\mathcal{A}, N, B, \theta)$ where $\mathcal{A}$ is a finite set of agents, $N \subseteq \mathcal{A} \times \mathcal{A}$ a irreflexive and symmetric network, $B \subseteq \mathcal{A}$ a behavior, and $\theta \in [0, 1]$ the adoption threshold.[1] For an agent $a \in \mathcal{A}$, her neighborhood is $N(a) := \{b : (a, b) \in N\}$.

Threshold Model Dynamics. Threshold models are used to investigate the spread of a behavior over discrete time-steps $t_0, t_1, ...$, i.e., the dynamics of the

[1] The literature contains several variations, including infinite networks [16], non-inflating behavior [16], agent-specific thresholds, non-symmetric relations, weighted links [14], and multiple behaviors [1].

behavior. Given an initial threshold model for t_0, $\mathcal{M} = (\mathcal{A}, N, B_0, \theta)$, several update policies for the behavior set B_0 exists.[2] One popular such [7,10,14] is captured by (1):

$$B_{n+1} = B_n \cup \left\{ a : \frac{|N(a) \cap B_n|}{|N(a)|} \geq \theta \right\}. \tag{1}$$

I.e., a plays (adopts, follows) B at t_{n+1} iff a plays B at t_n, **or** a proportion of a's neighbors *larger or equal* to the threshold plays B at t_n.

The former disjunct makes B *increase* over time, i.e., $\forall n : B_n \subseteq B_{n+1}$. This guarantees that (1) reaches a fixed point. The 'or equal to' embeds a tie-breaking rule favoring B.

Inflation may be dropped and the tie-breaking rule changed by using e.g. the policy specified by (2) instead:

$$B_{n+1} = \left\{ a : \tfrac{|N(a)\cap B_n|}{|N(a)|} > \theta \right\} \cup \left\{ a : \tfrac{|N(a)\cap B_n|}{|N(a)|} = \theta \text{ and } a \in B_n \right\}. \tag{2}$$

The second set in the union invokes a conservative tie-breaking rule: if $\frac{|N(a)\cap B_n|}{|N(a)|} = \theta$, a will continue her behavior from t_n. That (2) does not cause B to inflate implies the possibility of *loops* in behavior, i.e. where $B_n = B_{n+2} \neq B_{n+1}$. Thereby (2) does not necessarily reach a fixed point.

Threshold Model Dynamics as Induced by Game Play. Threshold influence may naturally be seen as an instance of a coordination problem: given enough of an agent's neighbors adopt behavior B, the agent will seek to coordinate with that group by adopting B herself. This coordination problem may be modeled as a coordination game

	B	$\neg B$
B	x, x	$0, 0$
$\neg B$	$0, 0$	y, y

played on the network: at each time-step, each agent chooses one strategy from $\{B, \neg B\}$ and plays this strategy against all their neighbors simultaneously. Agent a's payoff t_n is then the sum of the payoffs of the $|N(a)|$ coordination games that a plays at time t_n. With these rules, B is a best-response for agent a at time t_n iff

$$x \cdot \tfrac{|N(a)\cap B_n|}{|N(a)|} \geq y \cdot \tfrac{|N(a)\cap \neg B_n|}{|N(a)|} \Leftrightarrow \tfrac{|N(a)\cap B_n|}{|N(a)|} \geq \frac{y}{x+y}. \tag{3}$$

Setting $\theta := \frac{y}{x+y}$, the right-hand side of (3) resembles the specifications from (1) and (2). The precise correlation is that (2) captures the best-response dynamics for coordination game play on networks when using conservative tie-breaking [16], while (1) captures the same with tie-breaking biased towards B and the added assumption of a (possibly irrational) 'seed' of agents always playing B [10].

[2] Attention is here restricted to deterministic, discrete time simultaneous updates. See e.g. [17] for stochastic processes.

2 Threshold Models, Kripke Models and Action Models

A threshold model gives rise to a Kripke model [5] with $\mathcal{A}$ as domain, N as relation and a valuation $\|\cdot\| : \Phi \longrightarrow \mathcal{P}(\mathcal{A})$, $\Phi := \{B\}$, determining the extension of the B playing agents. To describe features of agents' neighborhoods, we use a language $\mathcal{L}$ with suitable *threshold modalities*:

$$\top \mid B \mid \neg\phi \mid \phi \wedge \phi \mid \langle\leq\rangle\phi \mid [\leq]\phi \mid (=)\phi$$

The three operators could be parametrized by θ, but to lighten notation, we leave the threshold implicit.

Intuitively, if a satisfies $\langle\leq\rangle\phi$, then there exists a θ 'large enough' set of a's neighbors that satisfy ϕ. E.g., if $\phi := B$, then at least a θ fraction of a's neighbors satisfy B. According to (1), a should then change his behavior to B. The operator is inspired by [2,13] and exemplified in Fig. 2. $[\leq]$ is the universal 'box' to the existential 'diamond' $\langle\leq\rangle$: if a satisfies $[\leq]\phi$, then all neighbors in all θ 'large enough' subsets of a's neighborhood satisfy ϕ. Finally, $(=)\phi$ captures that exactly a θ fraction of the agent's neighbors satisfy ϕ. In particular, if a satisfies $(=)B$, then a should invoke a tie-breaking rule.

With threshold θ, satisfaction in $\mathcal{M}$ is given by standard Boolean clauses and the following:

$$\begin{array}{lll}
\mathcal{M}, a \models B & \text{iff} & a \in \|B\| \\
\mathcal{M}, a \models \langle\leq\rangle\phi & \text{iff} & \exists C : \theta \leq \frac{|C \cap N(a)|}{|N(a)|} \text{ and } \forall a \in C,\ \mathcal{M}, a \models \phi \\
\mathcal{M}, a \models [\leq]\phi & \text{iff} & \forall C : \theta \leq \frac{|C \cap N(a)|}{|N(a)|} \text{ implies } \forall a \in C,\ \mathcal{M}, a \models \phi \\
\mathcal{M}, a \models (=)\phi & \text{iff} & \theta = \frac{N(a) \cap \|\phi\|_{\mathcal{M}}}{|N(a)|}.
\end{array}$$

The extension of ϕ in $\mathcal{M}$ is denoted $\|\phi\|_{\mathcal{M}} := \{a \in \mathcal{A} : \mathcal{M}, a \models \phi\}$.

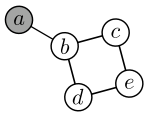


Fig. 2. A threshold model $\mathcal{M}$ with $\theta = \frac{1}{4}$ and B marked by gray. b satisfies $\langle\leq\rangle B$, as $\mathcal{M}, a \models B$ and $\frac{|\{a\}|}{|\{a,b,c\}|} \geq \frac{1}{4}$. Agent e satisfies $[\leq]\neg B$ as $\forall C \subseteq N(e) : \frac{|C \cap N(e)|}{|N(e)|} \geq \theta$ (that is, for sets $\{c\}, \{d\}, \{c,d\}$), $C \subseteq \|\neg B\|_{\mathcal{M}}$. Moreover, agent a satisfies $\neg(=)B \wedge [\leq]\neg B$ – hence, according to (2), she then should start playing $\neg B$, whereas (1) will not allow her to change.

From $\langle\leq\rangle$, $[\leq]$ and $(=)$, we define strict versions of the two former. These are useful when encoding non-biased tie-breaking rules:

$$\langle<\rangle\phi := \langle\leq\rangle\phi \wedge \neg(=)\phi$$
$$[<]\phi := [\leq]\phi \wedge \neg(=)\phi$$

Two comments on the threshold operators are due. First, the operators do not form the basis of a *normal* modal logic: $(=)$ distributes over neither $\vee$ nor $\wedge$, and the 'diamond' $\langle\leq\rangle$ does not distribute over $\vee$.[3] The 'box' $[\leq]$ does validate **K**: $[\leq](\phi \to \psi) \to ([\leq]\phi \to [\leq]\psi)$ and thus distributes over $\wedge$, but it is not the dual of $\langle\leq\rangle$, i.e., $[\leq]\phi \leftrightarrow \neg\langle\leq\rangle\neg\phi$ is not valid.[4] If $|\mathcal{A}| > 1$, the right-to-left direction holds, but not vice versa. Second, $(=)\phi$ does not imply that $(=)\neg\phi$, as the semantics are given w.r.t. θ. $(=)\phi$ does imply that $\frac{N(a)\cap\|\neg\phi\|_{\mathcal{M}}}{|N(a)|} = 1-\theta$. This point is important as only $\mathcal{M}, a \models (=)B$, and not $\mathcal{M}, a \models (=)\neg B$, means that a must invoke a tie-breaking rule.

Action Models and Product Update. Rather than updating threshold models by analyzing best responses or consulting equations like (1) or (2), they may be updated by taking the graph-theoretical product with a graph that encodes *decision rules*, uniformly followed by all agents. Such graphs are known as *action models (with postconditions)* [3,4,9]. To illustrate, then (cf. Proposition 1 below) $\mathcal{E}_1$ captures the same dynamics as those invoked by (1):

$\mathcal{E}_1$: σ_1: $(\langle\leq\rangle B, B)$ — σ_2: $(\neg\langle\leq\rangle B, \top)$

In the current context, it is natural to interpret each state of an action models as a decision rule.[5] E.g., σ_1 encodes the rule 'if a θ fraction or more of your neighbors play B, then play B'. State σ_2 encodes that if the agent is not influenced to play B, she should continue her current behavior.

Formally, by an *action model* we here refer to a tuple $\mathcal{E} = (|\mathcal{E}|, R, cond)$ where $|\mathcal{E}|$ is a non-empty domain of states, $R \subseteq |\mathcal{E}| \times |\mathcal{E}|$ is a relation on $|\mathcal{E}|$, and *cond* a pre- and postcondition map $cond : |\mathcal{E}| \longrightarrow \mathcal{L} \times \{B, \neg B, \top\}$ with $cond(\sigma) = (\phi, \psi) =: (pre(\sigma), post(\sigma))$.

The *product update* [3,9] of threshold model $\mathcal{M}$ and action model $\mathcal{E}$ is the threshold model $\mathcal{M} \otimes \mathcal{E} = (\mathcal{A}^{\uparrow}, N^{\uparrow}, B^{\uparrow}, \theta)$ with θ from $\mathcal{M}$, and

$$\mathcal{A}^{\uparrow} = \{(a, \sigma) \in \mathcal{A} \times |\mathcal{E}| : \mathcal{M}, a \models pre(\sigma)\},$$

$$N^{\uparrow} \ni ((a, \sigma), (b, \sigma')) \text{ iff } (a, b) \in N \text{ and } (\sigma, \sigma') \in R, \text{ and}$$

$$B^{\uparrow} = \{(s, \sigma) : s \in B \wedge post(\sigma) \neq \neg B\} \cup \{(s, \sigma) : post(\sigma) = B\}.$$

By the last condition, $B^{\uparrow}$ consists of 1) the agents in B minus those who change to $\neg B$, plus 2) the agents that change to B. Hence every agent will after the update again only play one strategy. If $post(\sigma) = \top$, no change in behavior is invoked.

[3] The latter was pointed out by Prof. A. Baltag for a similar operator in [2].

[4] The dual of $[\leq]$ would have the universal quantifier in the semantic clause of $\langle\leq\rangle\phi$ replaced by an existential one.

[5] The relation between actions is merely a technicality and is not given an interpretation. Given a re-defined product operation that ignores the relation of the action model, it could from both a conceptual and technical point be omitted.

3 Action Model Dynamics

Considering threshold models as Kripke models, it is possible to construct action models that when applied using product update will produce model sequences step-wise equivalent to those produced by (1) and (2). Moreover, the used models (in particular $\mathcal{E}_2$ below) gives rise to a simple class of action models. This class, specified below, contains all natural variations of the decision rules emulating (1) and (2). Thus, the class specifies all the different sets of decision rules by which agents may update their behavior while still behaving in the spirit of present notion of threshold influence.

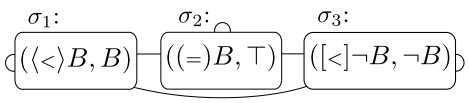
Proposition 1. *For any threshold model $\mathcal{M}$, the action model $\mathcal{E}_1$ applied using product update produces model sequences step-wise equivalent to those of (1).*

Proof. Let $\mathcal{M} = (\mathcal{A}, N, B, \theta)$ be arbitrary with (1)-update $(\mathcal{A}, N, B^+, \theta)$ and $\mathcal{E}_1$-update $(\mathcal{A}^\uparrow, N^\uparrow, B^\uparrow, \theta)$. Then $f{:}a \mapsto (a, \sigma), \sigma \in \{\sigma_1, \sigma_2\}$ is an isomorphism from $(\mathcal{A}, N, B^+)$ to $(\mathcal{A}^\uparrow, N^\uparrow, B^\uparrow)$. 1) $|\mathcal{A}| = |\mathcal{A}^\uparrow|$, as the preconditions of $\mathcal{E}_1$ partitions $\mathcal{A}$ entailing that no agents multiply or die under product update. 2) $((a, \sigma), (b, \sigma')) \in N^\uparrow$ iff $(a, b) \in N$: R from $\mathcal{E}_1$ is the full relation, so N dictates $N^\uparrow$. 3) $f(B^+) = B^\uparrow$ as

$$
\begin{array}{rll}
a \in B^+ & \textbf{iff} & \\
\frac{|N(a) \cap B|}{|N(a)|} \geq \theta & \textbf{iff} & \\
\exists C \subseteq N(a) \cap B : \frac{|C|}{|N(a)|} \geq \theta & \textbf{iff} & \\
\mathcal{M}, a \models \langle \leq \rangle B & \textbf{iff} & \\
\mathcal{M}, a \models pre(\sigma_1) & \textbf{iff} & \\
\mathcal{M} \otimes \mathcal{E}_1, (a, \sigma_1) \models B & \textbf{iff} & f(a) \in B^\uparrow
\end{array}
$$

□

The action model $\mathcal{E}_1$ contains only two states as (1) invokes a biased tie-breaking rule, subsumed in the state σ_1 by using the non-strict $\langle \leq \rangle B$ in the precondition. (2), in contrast, invokes a conservative, unbiased tie-breaking rule. This requires an extra state to encode:



Interpreted as decision rules, σ_1 of $\mathcal{E}_2$ states that if strictly more than a θ fraction of an agent a's neighbors plays B, then a should do the same; σ_2 embodies the conservative tie-breaking rule: if *exactly* a θ fraction of a's neighbors play B (and hence a $(1-\theta)$ fraction plays $\neg B$), then a should not change her behavior; finally, for σ_3, notice that if $[<]\neg B$, i.e., that all θ 'strictly large enough' subsets of a's neighbors plays $\neg B$, then there is a strictly larger than $(1-\theta)$ fraction of

her neighbors that play $\neg B$—σ_3 states that in that case, a should also play $\neg B$.

Proposition 2. *For any threshold model $\mathcal{M}$, The action model $\mathcal{E}_2$ applied using product update produces model sequences step-wise equivalent to those of (2).*

Proof. Analogous to those of Propositions 1 and 3 (see below). □

The Class of Threshold Model Update Action Models. For the reasons mentioned in the proof of Proposition 1, for an action model to change neither agent set nor network when applied to an arbitrary threshold model, it must be fully connected and it's preconditions must form a partition on the agent set. If one further accepts only preconditions that are in the spirit of standard threshold model updates, i.e., that agents change behavior based only on the behavior of their immediate neighbors, then the class of 'threshold model update action models' is easy to map. For by the latter restriction, $\langle < \rangle B$, $(=)B$ and $[>]\neg B$ form the unique finest partition[6] on the agent set of any threshold model. Given the three possible postconditions $B, \top$ and $\neg B$, the class of suitable action models contains 27 models (Table 1).

pre:	1	2	3	4	5	6	7	8	9
σ_1: $\langle < \rangle B$	B	B	B	B	B	B	B	B	B
σ_2: $(=)B$	B	B	B	$\top$	$\top$	$\top$	$\neg B$	$\neg B$	$\neg B$
σ_3: $[>]\neg B$	B	$\top$	$\neg B$	B	$\top$	$\neg B$	B	$\top$	$\neg B$

	10	11	12	13	14	15	16	17	18
σ_1: $\langle < \rangle B$	$\top$	$\top$	$\top$	$\top$	$\top$	$\top$	$\top$	$\top$	$\top$
σ_2: $(=)B$	B	B	B	$\top$	$\top$	$\top$	$\neg B$	$\neg B$	$\neg B$
σ_3: $[>]\neg B$	B	$\top$	$\neg B$	B	$\top$	$\neg B$	B	$\top$	$\neg B$

	19	20	21	22	23	24	25	26	27
σ_1: $\langle < \rangle B$	$\neg B$	$\neg B$	$\neg B$	$\neg B$	$\neg B$	$\neg B$	$\neg B$	$\neg B$	$\neg B$
σ_2: $(=)B$	B	B	B	$\top$	$\top$	$\top$	$\neg B$	$\neg B$	$\neg B$
σ_3: $[>]\neg B$	B	$\top$	$\neg B$	B	$\top$	$\neg B$	B	$\top$	$\neg B$

Table 1. Each action model contains three states with preconditions specified by *pre* and postconditions by columns 1 to 27.

As mention, this class of action models may be seen as containing all the possible sets of decision rules compatible with the used notion of threshold influence. Using action models it is a simple, combinatorial task to map. This is a benefit of using action models to define dynamics over the set theoretic specification.

[6] The symmetric variant $\langle < \rangle \neg B$, $(=)\neg B$ and $[>]B$ is ignored as it is equivalent up to interchange of B and $\neg B$.

Dynamics Induced by Action Models. Note that the action model $\mathcal{E}_1$ is not explicitly listed in Table 1. It is not so as $\mathcal{E}_1$ is based on a coarser partition of the agent set, containing two rather than three cells. It is however equivalent to the listed model 2: simply collapse states σ_1 and σ_2 to one.

The class include three trivial dynamics induced by models 1, 14 and 27 and seven that make little sense (4, 7, 8, 16, 17 and 24).

The best-response dynamics of coordination games are emulated by models 3, 6 and 9, capturing discriminating (3,9) and conservative (6) tie-breaking (cf. Proposition 2), while models 2, 5, 15 and 18 capture inflating ('seeded') coordination game dynamics.

Proposition 3 below lends credences to the conjecture that models 19, 22 and 25 capture the best-response dynamics for anti-coordination games with discriminating (19,25) and conservative (22) tie-breaking, and that 10, 13, 23 and 26 capture inflating dynamics of anti-coordination games.

Proposition 3. *For any threshold model, the best-response dynamics of the anti-coordination game*

	B	$\neg B$
B	$0, 0$	y, x
$\neg B$	x, y	$0, 0$

played with the conservative tie-breaking rule is step-wise equivalent to applying the action model 22 ($\mathcal{E}_{22}$) from Table 1 with $\theta = \frac{x}{y+x}$.

Proof. Let $\mathcal{M} = (\mathcal{A}, N, B, \theta)$. Playing B is a best-response in $\mathcal{M}$ for agent a iff

$$y \cdot \tfrac{|N(a)\cap\neg B|}{|N(a)|} \geq x \cdot \tfrac{|N(a)\cap B|}{|N(a)|} \Leftrightarrow \tfrac{|N(a)\cap\neg B|}{|N(a)|} \geq \tfrac{x}{y+x} = \theta$$

Hence, given the tie-breaking rule, the next set of B-players will be

$$B^+ = \{a : \tfrac{|N(a)\cap\neg B|}{|N(a)|} > \theta\} \cup \{a : \tfrac{|N(a)\cap\neg B|}{|N(a)|} = \theta \text{ and } a \in B\}.$$

Let $\mathcal{M} \otimes \mathcal{E}_{22} = (\mathcal{A}^\uparrow, N^\uparrow, B_n^\uparrow, \theta)$. Then $g : a \mapsto (a, \sigma)$, $\sigma \in |\mathcal{E}_{22}|$, is an isomorphism from $(\mathcal{A}, N, B^+)$ to $(\mathcal{A}^\uparrow, N^\uparrow, B_n^\uparrow)$. That $(\mathcal{A}, N) \cong_g (\mathcal{A}^\uparrow, N^\uparrow)$ follows from the proof of Proposition 1.

$$\begin{array}{ccccc}
a \in B^+ \textbf{ iff} & & & & \\
\tfrac{|N(a)\cap\neg B|}{|N(a)|} > \theta & \text{or} & \tfrac{|N(a)\cap\neg B|}{|N(a)|} = \theta \text{ and } a \in B & \textbf{iff} & \\
\mathcal{M}, a \models \langle< \rangle \neg B & \text{or} & \mathcal{M}, a \models B \wedge (=)\neg B & \textbf{iff} & \\
\mathcal{M}, a \models pre(\sigma_1) & \text{or} & \mathcal{M}, a \models B \wedge pre(\sigma_2) & \textbf{iff} & \\
\mathcal{M} \otimes \mathcal{E}_{22}, (a, \sigma_1) \models B & \text{or} & \mathcal{M} \otimes \mathcal{E}_{22}, (a, \sigma_2) \models B & \textbf{iff } g(a) \in B^\uparrow & \square \\
(\text{as } post(\sigma_1) = B) & & (\text{as } post(\sigma_2) = \top) & &
\end{array}$$

Logics for Threshold Dynamics. Given the uniform, action model approach to the dynamics outlined, it may be conjectured that the dynamics may also be treated by a uniform logical approach, particularly the reduction axiom method well-known from dynamic epistemic logic [3,8].

Three things are required to obtain a complete logic for one of the dynamics:

(1) A complete axiomatization for the threshold operators $\langle\leq\rangle, [\leq]$ and $(=)$,
(2) A complete axiomatization of the network properties, and
(3) Reduction laws for the used action model.

For 1, one may search for results in the literature on probabilistic modal logic. No suitable, general result is known to the author. 2 is easily obtained, though it requires a richer language, extending $\mathcal{L}$ with a normal modal operator $\Diamond$ and hybrid logical *nominals* $\{i, j, ...\}$. The latter is required to express the irreflexivity of the network relation, characterized by $i \to \neg\Diamond i$. To complete a combined logic, interaction axioms for the thresholds operators and normal modal operators should also be added. A reduction axiom-based logic for action models with post-conditions already exists (the logic **UM** from [9]), but the system should be modified to suit the hybrid nominals and threshold modalities. If such a combined logic is obtained for one of the dynamics, one will automatically obtain complete logics for all of the 27 dynamics induced by the action models of Table 1, with the only variation between them being the used action model in the dynamic modalities.

4 An Action Model for 'Belief Change in the Community'

One reviewer asked whether there is a relation between the action model approach used here, and the finite state automata approach introduced in [24] for threshold influence dynamics of preferences, and in particular, whether a translation between the two approaches exist. We conjecture that this is indeed the case. To lend credence to this conjecture, we show this may be done for the slightly simpler framework of threshold influence of belief change from [15].

The basic framework of [15] investigates the dynamics of *strong* and *weak influence* of beliefs among agents in a symmetric and irreflexive network. Beliefs are represented by three mutually exclusive atoms $Bp, B\neg p$ and Up, evaluated at agents in the network, as above. $\mathcal{M}, a \models Bp$ reads 'a believes p', and being undecided about p, Up, is equivalent to $\neg Bp \wedge \neg B\neg p$. To describe the network, a normal box operator F is used: $\mathcal{M}, a \models F\phi$ iff $\forall b \in N(a), \mathcal{M}, b \models \phi$. F has dual $\langle F\rangle$ – $\langle F\rangle\phi$ reads 'I have a friend that satisfies ϕ'. Call the language $\mathcal{L}'$.

An agent is strongly influence to believe $\phi \in \{p, \neg p\}$ if all her friends believe ϕ, and weakly influenced to believe ϕ if no friends believe $\neg\phi$ while at least one friend believes ϕ. With

$$S\phi := FB\phi \wedge \langle F\rangle B\phi \text{ and}$$
$$W\phi := F\neg B\neg\phi \wedge \langle F\rangle B\phi,$$

the dynamics of strong and weak influence are then characterized by the finite state automaton in Fig. 3, applied to all agents simultaneously.

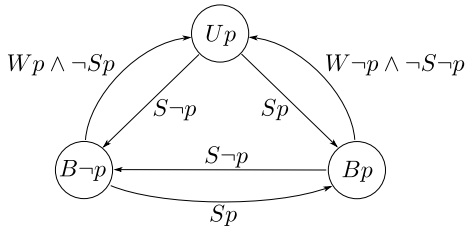


Fig. 3. The automaton of [15], which characterizes agents' belief change under weak and strong influence. If an agent is undecided about p, i.e., in the state Up, and is strongly influenced to believe p, Sp, she will change to state Bp, i.e., believe p. The automaton is deterministic.

Given this setup, it is no hard task to construct an action model over $\mathcal{L}'$ that will invoke the same dynamics. This may be done systematically by the construction: 1) for each state-transition-state triple (s, t, s') from the automaton, construct an action model state σ with the conjunction of the labels of s and t as precondition, and the label from s' as postcondition, and 2) let the relation of the action model be the full relation. The resulting action model $\mathcal{I}$ is depicted in Fig. 4. It is easy to verify that the effects of the two approaches are equivalent.

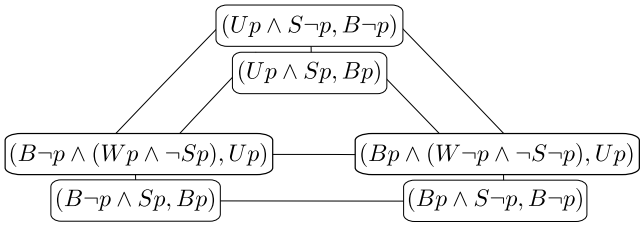


Fig. 4. The action model $\mathcal{I}$, invoking the same dynamics as the automaton of Fig. 3 (some edges are omitted). The top-most state makes an agent change from state Up to state $B\neg p$ if the agents also is strongly influenced to believe $\neg p$, etc.

The construction method used defines a function from automata to action models. If one restricts attention to action models with preconditions of the form $(\phi \wedge \psi)$, a function from action models to automata may be defined by the construction: 1) for each action model state σ, $cond(\sigma) = ((\phi \wedge \psi), \chi)$, construct a automaton state with label ϕ and one with label χ, and collapse all automata states with equivalent labels, and 2) for each all automaton states with labels ϕ and χ, add a transition with label ψ between them if there exists an action model state with $cond(\sigma) = ((\phi \wedge \psi), \chi)$. Combining the two constructions provides a bijection, serving as translation.

A Logic for Belief Change in the Community. Given that it is possible to emulate the dynamics invoked by the finite state automaton using an action

model, finding a sound and complete logic for the dynamics should be unproblematic. In fact, as F is a normal modal operator, the case is simpler than for threshold dynamics. Again some hybrid machinery is required to capture the irreflexive frame condition, but if this requirement is dropped, the reduction axiom system from [9] provides the desired result.

5 Closing Remarks

It has been argued that action models may be used to emulate the best-response dynamics on coordination and anti-coordination games played on networks by showing the product updates equivalent to the threshold model dynamics induced by game play, and that the method is applicable to the framework of threshold influence from [15]. It is conjectured that the action model approach to threshold dynamics lightens the work of finding complete logics, using methods well-known from dynamic epistemic logic, hereby providing new connections between game theory, social network theory and dynamic 'epistemic' logic.

Two questions present themselves. First, is it possible to rationalize the seven unaccounted for action models in the identified class, by moving from action models to game playing situations? Second, what is the extent of the applicability of action models? The present paper utilizes only a fraction of the potential of action models, as such may also be used to systematically alter the agent set and network. Changing the agent set may be used to model agent death and birth, whereby deterministic SIRS-like epidemiological dynamics [17] may be captured. Alterations to the social network may be used to model e.g. rise in popularity of information sources.

References

1. Apt, K., Markakis, E.: Diffusion in Social Networks with Competing Products. In: Persiano, G. (ed.) SAGT 2011. pp. 212–223. LNCS 6982, Springer (2011)
2. Baltag, A.: Modal Logics for Social Networks (Talk). Tsinghua Meets the ILLC Tsinghua University (2013)
3. Baltag, A., Moss, L.S., Solecki, S.: The Logic of Public Announcements, Common Knowledge, and Private Suspicions (extended abstract). In: Proc. of the intl. conf. TARK 1998. pp. 43–56. Morgan Kaufmann Publishers (1998)
4. van Benthem, J., van Eijck, J., Kooi, B.: Logics of communication and change. Information and Computation 204(11), 1620–1662 (2006)
5. Blackburn, P., de Rijke, M., Venema, Y.: Modal Logic. Cambridge University Press (2001)
6. Christoff, Z., Hansen, J.U.: A two-tiered formalization of social influence. In: Grossi, D., Roy, O., Huang, H. (eds.) Logic, Rationality, and Interaction, pp. 68–81. LNCS 8196, Springer (2013)
7. Christoff, Z., Rendsvig, R.K.: Dynamic logics for threshold models and their epistemic extension. ELISIEM, ESSLLI 2014 (2014)
8. van Ditmarsch, H., van der Hoek, W., Kooi, B.: Dynamic Epistemic Logic. Springer (2008)
9. van Ditmarsch, H., Kooi, B.: Semantic Results for Ontic and Epistemic Change. In: Bonanno, G., van der Hoek, W., Wooldridge, M. (eds.) Logic and the Foundations of Game and Decision Theory (LOFT 7). pp. 87–117. No. Loft 7 in Texts in Logic and Games, Vol. 3, Amsterdam University Press (2008)
10. Easley, D., Kleinberg, J.: Networks, Crowds, and Markets. Cambridge University Press (2010)
11. Granovetter, M.: Threshold Models of Collective Behavior. American Journal of Sociology 83(6), 1420–1443 (1978)
12. Hansen, P.G., Hendricks, V.F., Rendsvig, R.K.: Infostorms. Metaphilosophy 44(3), 301–326 (2013)
13. Heifetz, A., Mongin, P.: The Modal Logic of Probability. In: TARK '98 Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge. pp. 175–185. Morgan Kaufmann Publishers (1998)
14. Kempe, D., Kleinberg, J., Tardos, E.: Maximizing the Spread of Influence through a Social Network. In: Proc. 9th ACM SIGKDD intl. conf. on Knowledge Discovery and Data Mining, pp. 137–146. ACM, New York, NY, USA (2003)
15. Liu, F., Seligman, J., Girard, P.: Logical dynamics of belief change in the community. Synthese 191(11), 2403–2431 (2014)
16. Morris, S.: Contagion. Review of Economic Studies 67, 57–78 (2000)
17. Newman, M.E.J.: The spread of epidemic disease on networks. Physical Review E 66(1), 016128 (2002)
18. Ruan, J., Thielscher, M.: A logic for knowledge flow in social networks. In: Wang, D., Reynolds, M. (eds.) AI 2011: Advances in Artificial Intelligence, Lecture Notes in Computer Science, vol. 7106, pp. 511–520. Springer Berlin Heidelberg (2011)
19. Schelling, T.: Micromotives and Macrobehavior. Norton (1978)
20. Seligman, J., Liu, F., Girard, P.: Logic in the community. In: Banerjee, M., Seth, A. (eds.) Logic and Its Applications, LNCS, vol. 6521, pp. 178–188. Springer (2011)
21. Shakarian, P., Eyre, S., Paulo, D.: A Scalable Heuristic for Viral Marketing Under the Tipping Model. Social Network Analysis and Mining 3(4), 1225–1248 (2013)

22. Skyrms, B.: Evolution of the Social Contract. Cambridge University Press, Cambridge (1996)
23. Valente, T.: Social network thresholds in the diffusion of innovations. Social Networks 18(1), 69–89 (Jan 1996)
24. Zhen, L., Seligman, J.: A logical model of the dynamics of peer pressure. Electronic Notes in Theoretical Computer Science 278, 275 – 288 (2011)

Hierarchy of Expressive Power in Public Announcement Logic with Common Knowledge

Fangzhou Zhai and Tingxiang Zou

ILLC, University of Amsterdam

Abstract In this paper we investigate the expressive power of some fragments of Public Announcement Logic with Common Knowledge (PALC) on Kripke models. We will see that constrains on the formula that can be announced yields a hierarchy of expressive power, both on general Kripke models, and on $\boldsymbol{S5}$ models.

Keywords: hierarchy of expressive power, game semantics, public announcement logic, common knowledge

1 Introduction

It is well established that publish announcement logic with common knowledge[4] operator(PAC) is strictly more expressive on Kripke models than its counterpart without common knowledge(PA)[1] ; also, announcements solely does not increase the expressive power of epistemic logic, i.e., PA is equally expressive with epistemic logic[3]. If we investigate into PAC, namely restrict the set of formulas that can be announced in the logic, we expect a hierarchy of expressive power as a consequence. In fact, if we consider some natural restrictions of the set of formulas that can be announced, i.e., only allow the announcement of formulas in propositional logic, epistemic logic, epistemic logic with common knowledge and PAC respectively, as we shall see, we will obtain a strict hierarchy of expressive power on general Kripke models; slightly differently, on $\boldsymbol{S5}$ models, most hierarchy stays the same, however, adding common knowledge operator to the announcements does not increase the expressive power of the logic.

2 Preliminaries

Firstly we define the fragments of PAC we are going to investigate.

Definition 1. *(Languages $\mathcal{L}_P, \mathcal{L}_E, \mathcal{L}_C, \mathcal{L}_{PA}$).*
Let **P** *be a set of atomic propositions and* **Agent** *be a set of agents, let languages*

$$\mathcal{L}_P : \phi ::= \top \mid p \mid \neg\phi \mid \phi \vee \phi$$

$$\mathcal{L}_E : \phi ::= \top \mid p \mid \neg\phi \mid \phi \vee \phi \mid \Box_a \phi$$

$$\mathcal{L}_C : \phi ::= \top \mid p \mid \neg\phi \mid \phi \vee \phi \mid \Box_a\phi \mid C\phi$$

where $p \in \mathbf{P}$, $a \in \mathbf{Agent}$.

The fragments $PAC \restriction_x$ *where* $x \in \{P, E, C\}$ *is defined as follows:*

$$\phi ::= \top \mid p \mid \neg\phi \mid \phi \vee \phi \mid \Box_\alpha\phi \mid C\phi \mid [!\psi]\phi$$

where $\psi \in \mathcal{L}_x$.

The modal depth of a formula ϕ is denoted by $d(\phi)$. In particular, we have $d([!\psi]\phi) = d(\psi) + d(\phi)$. We write

$$\mathcal{M}, w \equiv_n^x \mathcal{M}, w'$$

for $x \in \{P, E, C\}$ if the pointed models satisfy the same set of formulas up to modal depth n in the fragment $PAC \restriction_x$. We write

$$\mathcal{M}, w \equiv_n^{PAC} \mathcal{M}, w'$$

if the points satisfy the same set of formulas up to modal depth n in PAC itself.

Model comparison games will play a crucial role in our proofs. The model comparison game for PAC is given in [2]. By slightly modifying the definitions, we get the model comparison games for the fragments we are interested in.

Definition 2. *(The Model Comparison Games) Given two Kripke models* $\mathcal{M} = (W, R, V)$, $\mathcal{M}' = (W', R', V')$ *and* $w \in W$, $w' \in W'$, *the n-round model comparison game* $G_x(\mathcal{M}, w; \mathcal{M}', w')$, *where* $x \in \{P, E, C, PAC\}$ *is defined as follows. If* $n = 0$, *Spoiler wins if and only if* w *and* w' *differ in their valuations. Otherwise, in each round, spoiler can initiate one of the following scenarios.*

(1) $\Box_a$*-move: spoiler choose a point x in one model which is an a-successor of v or v', and duplicator responds by choosing a matching a-successor y in the other model. The move outputs x,y.*

(2) C-move: spoiler choose a point x in one model reachable by a path from current worlds v or v', and duplicator responds by choosing a matching point y reachable by a path in the other model. The move outputs x,y.

(3) $[\phi]$*-move: Spoiler choose a number* $r \leq n$ *and sets* $S \subseteq W$, $S' \subseteq W'$, *s.t.* $w \in S$ *and* $w' \in S'$. ***Stage 1****: Duplicator choose states* $s \in S \cup S'$, $\bar{s} \in \bar{S} \cup \bar{S'}$ *(where* $\bar{S}$ *is the complement of* S*). Then Spoiler and Duplicator play an r-round subgame with initial configuration s,* $\bar{s}$. *If Duplicator wins this subgame, he also wins the entire game.* ***Stage 2****: The configuration changes to* $M|S, w$ *and* $M'|S', w'$, *with* $n - r$ *rounds remaining.*

(4) If $x = P$, r *must be* 0. *If* $x = E$, *the r round game played in* ***Stage 1*** *does not contain* C_B*-moves nor* $[\phi]$*-moves. Similarly, if* $x = C$, *in* ***Stage 1****,* $[\phi]$*-moves are not allowed.)*

A player loses if he cannot perform a move. Spoiler wins if any matching worlds differ in their valuations. Duplicator wins otherwise. In particular, duplicator wins if the game does not terminate within finitely many moves(i.e., if spoiler keeps taking $[\phi]$ moves where $r = 0$).

Theorem 1. ***(Adequacy of model comparison games)*** *If the set of propositional letters* **P** *and agents* **Agent** *are both finite, duplicator has a winning strategy in the n-round game $G_x(\mathcal{M}, w; \mathcal{M}', w')$, where $x \in \{P, E, C, PAC\}$ if, and only if, $\mathcal{M}, w \equiv_n^x \mathcal{M}', w'$.*

In [2] the proof of adequacy of the model comparison game for the entire PAC has already been presented. By simplifying the proof, it is straightforward to obtain the proof of the adequacy of the model comparison games of the fragments we defined.

We now give the definition of expressive power.

Definition 3. ***(Expressive Power)*** *Given $\mathcal{L}_1 = \langle L_1, \mathcal{C}, \Vdash \rangle$, where L is the language, $\mathcal{C}$ is the class of models and $\Vdash$ is the semantics, and $\mathcal{L}_2 = \langle L_2, \mathcal{C}, \vDash \rangle$, we say that $\mathcal{L}_2$ is at least as expressive as $\mathcal{L}_1$, and write $\mathcal{L}_1 \leq_{\mathcal{C}} \mathcal{L}_2$, if for all $\phi_1 \in \mathcal{L}_1$ there exists $\phi_2 \in \mathcal{L}_2$ such that for all $\mathcal{M} \in \mathcal{C}$ we have $\mathcal{M} \Vdash \phi_1 \Leftrightarrow \mathcal{M} \vDash \phi_2$. The notions $=_{\mathcal{C}}$ and $<_{\mathcal{C}}$ are defined accordingly in the natural way.*

3 Main Results

In the first few subsections, we present the proofs of the following expressivity relations on $\boldsymbol{S5}$ models:

$$EL <_{\boldsymbol{S5}} PAC\restriction_P <_{\boldsymbol{S5}} PAC\restriction_E =_{\boldsymbol{S5}} PAC\restriction_C <_{\boldsymbol{S5}} PAC$$

The last subsection will be devoted to the expressivity relations on general Kripke models. Since strict inequality relations on $\boldsymbol{S5}$ models imply the same result on general Kripke models, given the results on $\boldsymbol{S5}$ models proved, what left to be investigated about expressive powers on the latter is that of $PAC\restriction_E$ and $PAC\restriction_C$. We shall see that, in contrast with the expressive power on $\boldsymbol{S5}$ models, we obtain another strict inequality. That would give us the following expressivity relations on general Kripke models:

$$EL <_{\boldsymbol{K}} PAC\restriction_P <_{\boldsymbol{K}} PAC\restriction_E <_{\boldsymbol{K}} PAC\restriction_C <_{\boldsymbol{K}} PAC$$

3.1 Epistemic Logic and PAC$\restriction_{\boldsymbol{P}}$

We now prove the following expressivity result.

Theorem 2. *$PAC\restriction_P$ is strictly more expressive than epistemic logic on $\boldsymbol{S5}$ models.*

We basically make use of the following lemma.

Lemma 1. *Two logics are equally expressive if, and only if, they distinguishes the same pairs of classes of models. i.e., given* $\mathcal{L}_1 = \langle L_1, \mathcal{C}, \Vdash \rangle$, $\mathcal{L}_2 = \langle L_2, \mathcal{C}, \vDash \rangle$, *we have* $\mathcal{L}_1 =_e \mathcal{L}_2$ *if for any* $\mathcal{C}_1, \mathcal{C}_2 \subseteq C$, $\mathcal{L}_1$ *distinguishes them, (i.e., there exists* $\phi_1 \in \mathcal{L}_1$ *such that* $\mathcal{M} \Vdash \phi_1$ *for all* $\mathcal{M} \in \mathcal{C}_1$ *and* $\mathbb{N} \nVdash \phi_2$ *for all* $\mathbb{N} \in \mathcal{C}_2$*) if, and only if,* $\mathcal{L}_2$ *distinguishes them.*

Proof is straightforward. The lemma indicates that if we can find classes of models C_1 and C_2 such that $PAC \restriction_P$ distinguishes the classes while epistemic logic does not, we will have the desired inequality result. We now define pointed **S5** models $\mathcal{M}_n, s_n$ and $\mathcal{M}_n, t_n$ such that some formula ϕ in $PAC \restriction_P$ distinguishes them while no formula in epistemic logic up to modal depth n does. Given that defined, it is clear that ϕ distinguishes the model classes $\mathcal{C}_1 = \{\mathcal{M}_n, s_n | n \in \omega\}$ and $\mathcal{C}_2 = \{\mathcal{M}_n, t_n | n \in \omega\}$ while no formula in epistemic logic does, thus finishing the proof. All the inequality results are proved using the same strategy.

We now prove the theorem.
Proof.
Consider the following models.
$\mathcal{M}_n = (M, R_a, R_b, R_c, V)$ where
$M = \{s_i, u_i, t_i, v_i | i \leq n\} \cup \{y\}$
R_a is the equivalent closure of
$\{(s_{2i}, s_{2i+1}), (s_{2i}, u_{2i}), (t_{2i}, t_{2i+1}), (t_{2i}, v_{2i}), (u_{2i+1}, v_{2i+1}) | i \leq n/2\}$
R_b is the equivalent closure of
$\{(s_{2i+1}, s_{2i+2}), (s_{2i+1}, u_{2i+1}), (t_{2i+1}, t_{2i+2}), (t_{2i+1}, v_{2i+1}), (u_{2i}, v_{2i}) | i \leq n/2\}$
R_c is the equivalent closure of
$\{(y, t_0)\}$
$V(p) = \{u_i, v_i | i \leq n\}$;
$V(q) = \{y\}$.
We denote the equivalent closure of $R_a \cup R_b \cup R_c$ by R_C.
The following graph illustrates the model $\mathcal{M}_4$. Note the the relations should be the equivalent closure of what is in the graph.

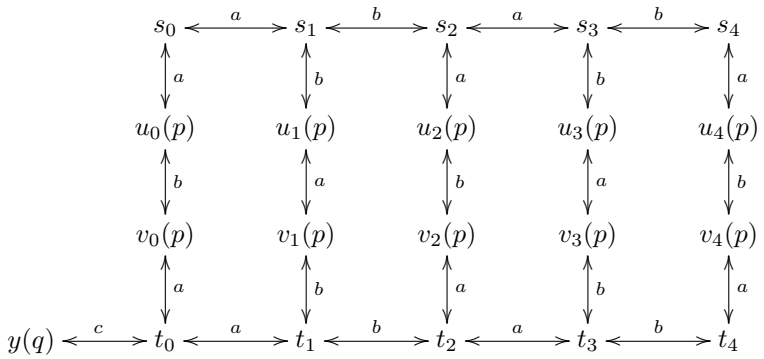


Figure 1. The model

Consider formula

$$\phi = [!\neg p]C\neg q$$

Clearly, the formula lies in $PAC \restriction_P$ but is not an epistemic logic formula. Clear still, $\mathcal{M}_n, s_n \models \phi$ while $\mathcal{M}_n, t_n \not\models \phi$. i.e., the fragment $PAC \restriction_P$ distinguishes the models classes $\mathcal{C}_1$ and C_2. We now show that no formula in epistemic logic up to modal depth n distinguishes models $\mathcal{M}_n, s_n$ and $\mathcal{M}_n, t_n$, which is quite straightforward: $\mathcal{M}_n, s_n$ and $\mathcal{M}_n, t_n$ are clearly n-bisimilar; moreover, they are connected to the same set of points in the model(i.e., the entire model) by the relation R_C. That means they satisfy the same set of formulas in epistemic logic with common knowledge up to modal depth n. As a consequence, no formula in epistemic logic with common knowledge distinguishes the classes $\mathcal{C}_1$ and $\mathcal{C}_2$. That finishes the proof. □

3.2 PAC$\restriction_P$ and PAC$\restriction_E$

We prove the following theorem using the same strategy as is used in the previous section.

Theorem 3. *$PAC \restriction_E$ is strictly more expressive than $PAC \restriction_P$ on **S5** models.*

Proof.
Consider the following models.
$\mathcal{M}_n = (M, R_a, R_b, R_c, R_d, V)$ where
$M = \{s_i, u_i, t_i, v_i, r_i | i \leq n\} \cup \{y\}$
R_a is the equivalent closure of
$\{(s_{2i}, s_{2i+1}), (s_{2i}, u_{2i}), (t_{2i}, t_{2i+1}), (t_{2i}, v_{2i}), (u_{2i+1}, v_{2i+1}) | i \leq n/2\}$
R_b is the equivalent closure of
$\{(s_{2i+1}, s_{2i+2}), (s_{2i+1}, u_{2i+1}), (t_{2i+1}, t_{2i+2}), (t_{2i+1}, v_{2i+1}), (u_{2i}, v_{2i}) | i \leq n/2\}$
R_c is the equivalent closure of
$\{(u_i, r_i), (v_i, r_i) | i \leq n\}$;
R_d is the equivalent closure of
$\{(y, t_0)\}$
$V(p) = \{r_i | i \leq n\}$;
$V(q) = \{y\}$.
The following figure illustrates the model $\mathcal{M}_3$. Still, the relations should be the equivalent closure of what is in the figure.

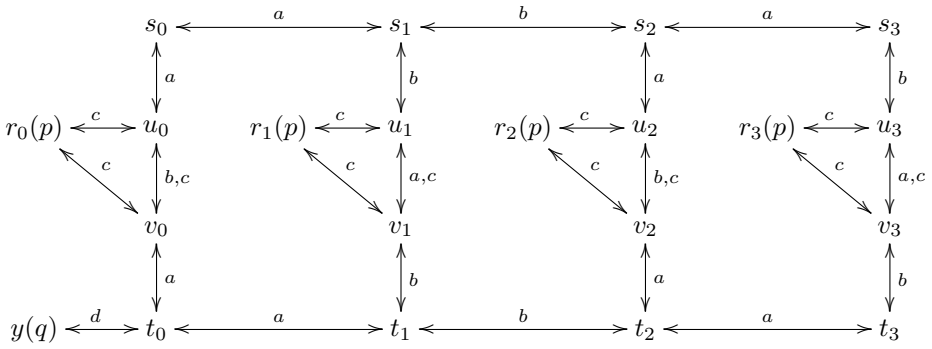


Figure 2. The model $\mathcal{M}_3$

Now consider the following formula, which lies in $PAC \restriction_E - PAC \restriction_P$.

$$\phi = [!\neg\Diamond_c p]C\neg q$$

The points $\{v_i, u_i, r_i | i \leq n\}$ are those that satisfies $\Diamond_c p$, and are deleted after the announcement. It is thus obvious that $\mathcal{M}_n, s_n \models \phi$ while $\mathcal{M}_n, t_n \not\models \phi$, i.e., ϕ distinguishes the classes $C_1 = \{\mathcal{M}_n, s_n | n \in \omega\}$ and $C_2 = \{\mathcal{M}_n, t_n | n \in \omega\}$. We now show that no formulas in $PAC \restriction_P$ up to modal depth n distinguishes the models $\mathcal{M}_n, s_n$ and $\mathcal{M}_n, t_n$, which, by previous argument, suffices to prove the theorem.

We give a winning strategy for duplicator in the n-round game $G_P(\mathcal{M}_n, s_n; \mathcal{M}_n, t_n)$. Observe that s_n and t_n are n-bisimilar. Therefore, if spoiler makes a $\Box_a$ move, duplicator can respond by taking a witness of bisimilarity; if the spoiler takes a C move, then, since s_n and t_n reaches the same set of points(i.e., the entire model), duplicator can choose the same point as a response, thus winning the game by following all the moves of the spoiler then on.

It is a little bit complicated in case spoiler takes a $[\phi]-$ move. Assume that spoiler choose sets $S, S' \subseteq M_n$. If $S \neq S'$, w.l.o.g., say we have $x \in S - S'$, we see that duplicator wins the 0-round subgame in **Stage 1** by choosing $x \in S$, and $x \in M_n - S'$(since they have the same valuation thus satisfy the same set of propositional formulas): spoiler loses the game immediately. Therefore spoiler must make sure that $S = S'$. By a similar argument, if points $x \in S$ and $x' \in M_n - S$ has the same valuation, spoiler also loses immediately. Now assume that no announcement has been made so far. If one point in the current configuration is y or some r_i, then by our definition of strategy, the two points that forms the configuration must be identical, therefore duplicator wins the game by following every move of spoiler. Otherwise, observe that points $\{s_i, t_i, v_i, u_i | i \leq n\}$ satisfies the same propositional formulas. Since it must be the case that the current configuration survives the relativization, so must all the rest of $\{s_i, t_i, v_i, u_i | i \leq n\}$. Now our arguments about the strategy still applies, and the strategy can be executed for n rounds. It is clearly winning by definition.

3.3 PAC$\restriction_E$ and PAC$\restriction_C$

In this section, we prove the following expressivity result.

Theorem 4. *$PAC \restriction_C$ is equally expressive as $PAC \restriction_E$ on **S5** models.*

Proof. The theorem tells us that common knowledge operator does not increase the expressive power of the fragment when added to the formulas that is allowed to be announced. This basically results from the following observation made true by the special property of **S5** models.

Observation 5. *For any formula $C\phi \in PAC$ and any **S5** model $\mathcal{M}$ that is connected under R_C, it is either the case that $M, w \models C\phi$ for all $w \in M$ or $\mathcal{M}, w \not\models C\phi$ for all $w \in M$.*

The observation indicates that an announcement $[!C\phi]$ is determined to be trivial: it is either going to delete all points in the model or preserve all points in the model. Now consider the following reduction axioms.

1. $[!C\phi]\psi \leftrightarrow C\phi \rightarrow \psi$
2. $[!\neg C\phi]\psi \leftrightarrow \neg C\phi \rightarrow \psi$
3. $[!\phi' \vee C\phi]\psi \leftrightarrow (C\phi \wedge \psi) \vee (\neg C\phi \wedge [!\phi']\psi)$
4. $[!\Box_a C\phi]\psi \leftrightarrow C\phi \rightarrow \psi$

Given the observation, it is straightforward to prove that these axioms are valid on **S5** models. In particular, if we consider a formula from $PAC \restriction_C$, these axioms allows us to move the common knowledge operator out of the announcement to obtain an equivalent formula in $PAC \restriction_E$. It is thus immediate that $PAC \restriction_C =_e PAC \restriction_E$.

3.4 PAC$\restriction_E$ and PAC

Theorem 6. *PAC is strictly more expressive than $PAC \restriction_E$ (thus than $PAC \restriction_C$) on **S5** models.*

Proof.
We use the same strategy as we have been using to prove inequality results. Consider the following models.
$\mathcal{M}_n = (M_n, R_a, R_b, R_c, R_d, V)$ where
$M_n = \{u_i, u_i', v_i, v_i', w_i, x_i, x_i', y_i, y_i', z_i | 1 \leq i \leq n\} \cup \{t_i, t_i' | i \leq 2\} \cup \{s, t_3, t_a, t_a', t_b, t_b'\}$
R_a is the equivalent closure of
$\{(s, u_1), (s, u_1'), (u_{2i}, u_{2i+1}), (u_{2i}', u_{2i+1}'), (t_1, v_1), (v_{2i}, v_{2i+1}), (t_1', v_1'), (v_{2i}', v_{2i+1}'),$
$(t_3, y_1), (y_{2i}, y_{2i+1}), (t_3, y_1'), (y_{2i}', y_{2i+1}'), (t_1, x_1), (x_{2i}, x_{2i+1}), (t_1', x_1'), (x_{2i}', x_{2i+1}')|$ for

all possible i}
R_b is the equivalent closure of
$\{(u_{2i+1}, u_{2i+2}), (u'_{2i+1}, u'_{2i+2}), (v_{2i+1}, v_{2i+2}), (v'_{2i+1}, v'_{2i+2}),$
$(y_{2i+1}, y_{2i+2}), (y'_{2i+1}, y'_{2i+2}), (x_{2i+1}, x_{2i+2}), (x'_{2i+1}, x'_{2i+2})|$for all possible i}
R_c is the equivalent closure of
$\{(u_n, t_0), (u'_n, t'_0), (v_n, t_0), (v'_n, t'_0), (x_n, t_b), (x'_n, t'_b), (y_n, t_2), (y'_n, t'_2), (t_1, w_1), (w_{2i}, w_{2i+1}),$
$(t_3, z_1), (z_{2i}, z_{2i+1})|$for all possible i}
R_d is the equivalent closure of
$\{(t_0, t_a), (t_2, t_b), (t'_0, t'_a), (t'_2, t'_b), (w_{2i+1}, w_{2i+2}), (z_{2i+1}, z_{2i+2})|$for all possible i}
$V(p) = \{w_n\}$;$V(q) = \{t_0, t'_0, t_2, t'_2, t_a, t'_a, t_b, t'_b\}$;$V(r) = \{z_n\}$.
Finally, $\mathbb{N}_n = \mathcal{M}_n \restriction_{N_n}$ where
N_n collects the points in M_n without a prime in its name.
The following figure shows the model $\mathcal{M}_1$. Still, the relationships should be the equivalent closure of what is shown in the figure.

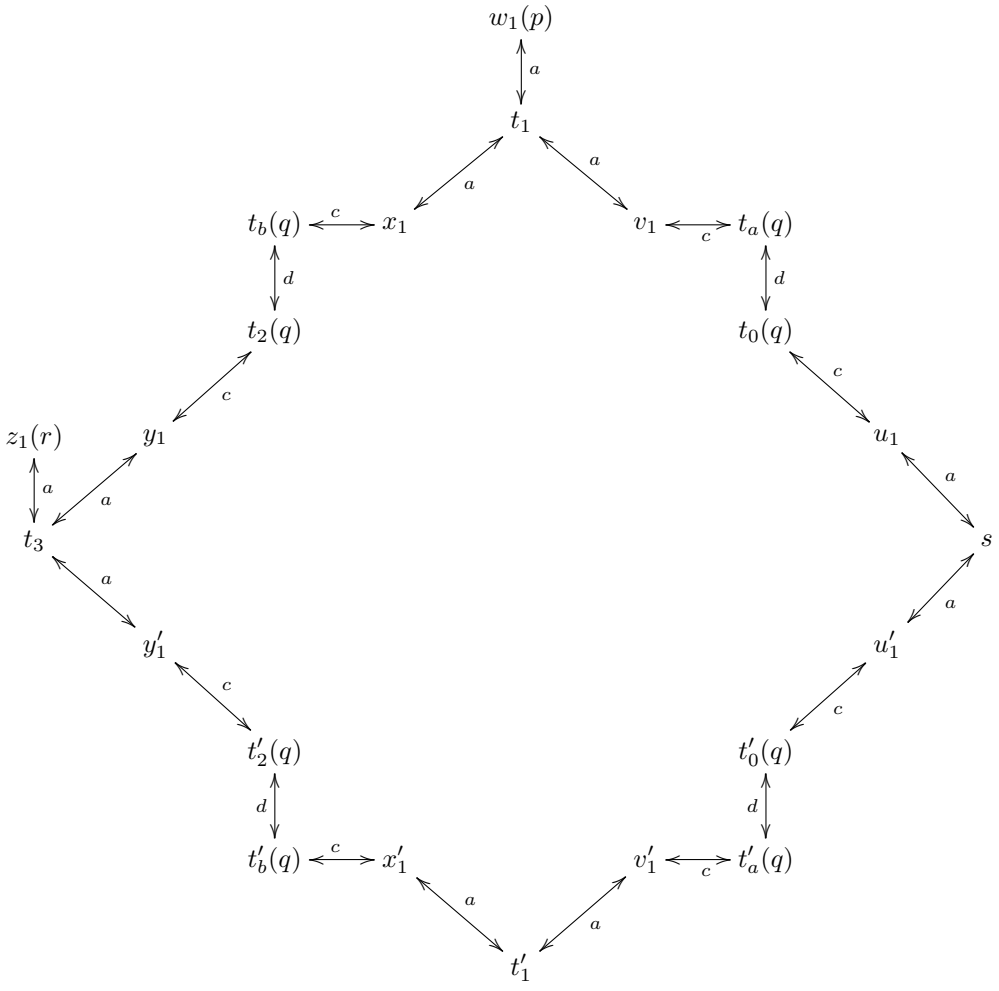


Figure 3. The model $\mathcal{M}_1$

The model $\mathbb{N}_1$ is simply $\mathcal{M}_1$ without the primed points.
Now consider formula

$$\phi = [![!\neg q]C\neg p]\neg C\neg r$$

Clearly the formula lies in $PAC - PAC\restriction_E$. After announcing $\neg q$, points $t_0, t_0', t_2, t_2', t_a, t_a', t_b, t_b'$ will be deleted. As a result, $C\neg p$ is only satisfied on points except $\{t_1, w_i, v_i, x_i | i \leq n\}$. Therefore after the announcement of $![!\neg q]C\neg p$, these points will be deleted. In the relativized model $\mathcal{M}_n'$(what remains of $\mathcal{M}_n$), points z_n is still reachable from s by the primed points thus $\mathcal{M}_n', s \models \neg C\neg r$ while in the relativezed model $\mathbb{N}_n'$, z_n is no longer reachable from s, therefore $\mathbb{N}_n, s \not\models \neg C\neg r$. As a summary, $\mathcal{M}_n, s \models \phi$ while $\mathbb{N}_n, s \not\models \phi$, i.e., $\phi \in PAC$ distinguishes $\mathcal{C}_1 = \{\mathcal{M}_n, s | n \geq 1\}$ and $\mathcal{C}_2 = \{\mathbb{N}_n, s | n \geq 1\}$.

We now prove that no formula in $PAC\restriction_E$ up to modal depth n distinguishes $\mathcal{M}_n, s$ and $\mathbb{N}_n, s$, which, by theorem 4, suffices to prove that $PAC >_e PAC\restriction_C$ on **S5** models. We do so by giving a winning strategy for duplicator in the n-round game $G_E(\mathcal{M}_n, s; \mathbb{N}_n, s)$.

In case spoiler makes a $\Box_a$ move, note that $\mathcal{M}_n, s$ and $\mathbb{N}_n, s$ are n-bisimilar. As long as the current configuration are r-bisimilar, where r is the number of rounds remaining, duplicator can respond with a witness of n-bisimilarity.

In case spoiler makes a C move, if spoiler chooses a point without prime, then duplicator choose the point with the same name; otherwise if spoiler chooses a primed point u_i' or y_i', duplicator chooses the counterpart without prime; if spoiler chooses v_i' or x_i', duplicator chooses u_i in the other model. Let r be the number of rounds remaining. If no announcement has been made, duplicator can always do so and by doing so, the new configuration remains r-bisimilar, making possible our strategy for $\Box_a$ moves.

In case spoiler makes a $[\phi]$ move and dose not lose in **Stage 1**(i.e., the set S given by spoiler yields a possible relativization by some formula in epistemic logic), assume that no announcement has been made so far, we need to consider the following possibilities.

1. Some point with one of the letters u,v,x and y in its name or from $\{t_0, t_0', t_2, t_2', t_a, t_a', t_b, t_b'\}$ is deleted.
Observe that points $P_i = \{u_i, u_i', v_i, v_i', x_i, x_i', y_i, y_i'\}$ are n-bisimilar(also $\{t_0, t_0', t_2, t_2', t_a, t_a', t_b, t_b'\}$), therefore if any one of them is deleted in the announcement, so is the rest of P_i: that cuts the model in to several disconnected parts. Observe that, what remains from points $\{v_i', x_i', t_1' | 1 \leq i \leq n\}$ is isomorphic to what remains from $\{u_i, s, u_i' | 1 \leq i \leq n\}$. Since duplicator has been following the strategy stated above, the configuration a, b of the game remains the same in the sense of isomorphism: there is an obvious isomorphism between the connected subgraphs

reachable from them that maps a to b, and as a consequence, duplicator has a winning strategy from now on.
2. If any point with letter w or z in its name is deleted(and **1.** does not happen), then either the configuration falls into some isomorphic fragments(what remains of $\{w_i|1 \leq i \leq n\}$ or $\{z_i|1 \leq i \leq n\}$ that contains the configuration) or the strategy we give for $\Box_a$ moves and $[\phi]$ moves are still executable, given that the n rounds have not been exhausted yet.
3. If one of s, t_1, t_1', t_3 is deleted, since they are (n-1)-bisimilar, either the n rounds are exhausted, thus duplicator wins immediately since the current configuration satisfies the same set of propositional formula, or they are all deleted, which lead us into a similar situation as is described in **1.**.

It is clear that the strategy is winning. □

3.5 Expressive powers on general Kripke models

As an immediate consequence of theorems 2, 3 and 6, we have the following result about the expressive powers of the fragments on general Kripke models.

Corollary 1. *The following expressivity results holds for the fragments on general Kripke models:*

$$EL <_{\boldsymbol{K}} PAC \restriction_P <_{\boldsymbol{K}} PAC \restriction_E$$

$$PAC \restriction_C <_{\boldsymbol{K}} PAC$$

Therefore it remains to investigate the expressive powers of $PAC \restriction_E$ and $PAC \restriction_C$ on general Kripke models. To use a similar method for an inequality result, consider the following models.

Definition 4. $\mathcal{M}_n = (M_n, R_n, V_n)$, *where*
$M_n = \{x_i \mid 0 \leq i \leq n\} \cup \{y_j \mid 0 \leq j \leq n\} \cup \{z_1, z_2, z_3, g\}$
$R_n = \{(x_i, x_{i-1}), (y_i, y_{i-1}) \mid 1 \leq i \leq n\} \cup \{(x_i, z_j), (y_i, z_j) \mid 0 \leq i \leq n, j = 1, 2\} \cup \{(z_2, x_n), (z_2, y_n), (z_1, z_1), (z_1, z_2), (z_2, z_3), (z_3, z_3), (y_0, g)\}$
$V_n(p) = M_n$
In case $n = 2$, *the model is as follows:*

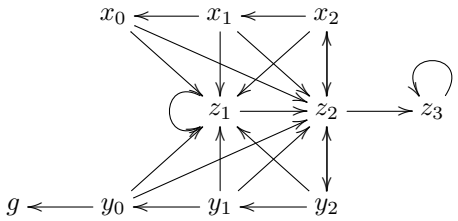


Figure 4. The model $\mathcal{M}_2$

Let

$$\varphi = [\neg \Diamond C \Diamond p] C \Diamond p$$

We see that φ lies in $PAC\restriction_C - PAC\restriction_E$. Furthermore, we have for any $n \in \mathbb{N}$, $\mathcal{M}_n, x_n \Vdash \varphi$ while $\mathcal{M}_n, y_n \nVdash \varphi$. Since z_3 is the only world that satisfies $\Diamond p$, after announcing $\neg\Diamond C\Diamond p$, z_2 and z_3 will be eliminated.
Another observation is that for all $0 \leq i \leq n$, each x_i and y_i are $i+1$ bisimilar. Similarly, y_i and z_j $(j = 1, 2, 3)$ are $i+1$ bisimilar.

Lemma 2. *For all $n \in \mathbb{N}$, duplicator has a winning strategy for the n-round game $G_E(M_n, x_n; M_n, y_n)$.*

Proof. Firstly we investigate what happens when spoiler take a $[\phi]$-move.
Assume that there are m rounds left and the configuration is now $M_n|S, x_i; M_n|S', y_i$. We assume w.l.o.g. that $S = S'$(otherwise duplicator has a winning strategy in Stage 1). Assume that spoiler chooses a number $0 \leq r \leq m$ and $x_i, y_i \in S'$. Then x_i, y_i, z_1, z_2, z_3 should be all in S', otherwise, since the points are bisimilar, duplicator will have a winning strategy in Stage 2. Then there are two cases for the relativized model.
In case that the path from y_i to g is not entirely contained in S', let S'' be the points reachable from x_i, then points in S'' are all bisimilar to each other: all the worlds in S'' can reach each other and they are consistent in atomic properties. Therefore in *Stage 2*, duplicator can choose arbitrary point to win.
Otherwise we have $S' = M_n$: since $y_r \in S'$ and all x_i, z_j, $i = 0, 1, \cdots, n$, $j = 1, 2, 3$ are r-bisimilar to each other, they are all in the relativised model(the update is trivial). Therefore it suffices to give a winning strategy of duplicator when spoiler takes a $\Box_a$ move or C move. However, if spoiler takes a C move, since y_n and z_n reaches the same set of points, duplicator can choose a same point to win the game; if spoiler takes a Box_a move and chooses z_i, duplicator can choose the same point; if spoiler choose y_{n-1}, duplicator can choose x_{n-1}(and vice versa), which yields a winning strategy inductively. As a summary, duplicator has a winning strategy. □

That gives us the following theorem.

Theorem 7. *$PAC\restriction_C$ is strictly more expressive than $PAC\restriction_E$.*

Therefore the claimed inequality result.

4 Conclusion and Future Work

We have proved the following expressivity results:

$$EL <_{\boldsymbol{K}} PAC\restriction_P <_{\boldsymbol{K}} PAC\restriction_E <_{\boldsymbol{K}} PAC\restriction_C <_{\boldsymbol{K}} PAC$$

$$EL <_{\boldsymbol{S5}} PAC\restriction_P <_{\boldsymbol{S5}} PAC\restriction_E =_{\boldsymbol{S5}} PAC\restriction_C <_{\boldsymbol{S5}} PAC$$

If we consider relative common knowledge operator instead, the conjecture is that the inequality results still holds. But the equality relation will probably fail since the key observation 5, which makes possible the reduction axioms, no longer holds. It is also interesting to investigate the expressivity of radical upgrades, or to investigate similar fragments of dynamic epistemic logic.

References

1. A. Baltag, L. S. Moss, S. Solecki: The logic of public announcements, common knowledge, and private suspicions. Technical Report. SEN-R9922 (1999)
2. Johan van Benthem, Jan van Eijck, Barteld Kooi: Logics of communication and change. Information and Computation. 204, 1620–1662(2006)
3. J.A.Plaza.: Logics of Public Communications. Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems. 201–206(1989)
4. Jon Barwise.: Three Views of Common Knowledge. Proceedings of the Second Conference on Theoretical Aspects of Reasoning About Knowledge. 365—379(1988).

Coherence and Extraction from Adjuncts in Chinese

Dawei Jin

Department of Linguistics, State University of New York at Buffalo,
609 Baldy Hall, 14260 Buffalo, USA
daweijin@buffalo.edu

Abstract The paper compares asymmetrical extraction from coordinate structures (in English) (that is, violations of the coordinate structure constraint) with asymmetrical extraction from subordinate structures (in Chinese) (that is violations of the adjunct island condition) and argues that these are grammatical if certain discourse requirements, in particular discourse relations (e.g. explanation, Occasion, common topichood etc) are met. These can be specific to the language and the kind of construction, depending also on the means of each language to express the discourse relations (by coordination or subordination).

Keywords: Coherence relations, adjunct sland condition, clause linkage, parasitic gaps, Chinese

1 Introduction

Empirical and experimental evidence over the years (Xu, 1990; Truswell, 2007; Heestand et al, 2011; Hofmeister et al, 2014) have shown that extraction from across adjuncts induces crosslinguistically less robust island effects, compared to extraction from across other strong island domains (such as complex NPs or subject clauses). This prompts debates as to whether the adjunct island effects follow from a syntactic locality constraint, as is commonly assumed for strong island effects in mainstream literature. My paper seeks to contribute to the debates by presenting an argument in favor of the nonsyntactic approaches. Specifically, I present new data to show that extraction from a structured subset of adjuncts in Chinese is sensitive to discourse-semantic coherence relations (Hobbs, 1979), which are originally invoked to predict extraction behaviors from conjuncts (Schmerling, 1972; Goldsmith, 1985; Lakoff, 1986; Deane, 1990; Na and Huck, 1992; Kehler, 2002).

This paper is organized as follows: in section 2 I summarize the coherence accounts of conjuncts; in section 3, I show that the extraction behaviors in Chinese temporal adjuncts strongly parallel those of English conjuncts; Section 4 argues that this parallelism can be accommodated by treating the conjunct-adjunct difference as following from an implicational semantics-syntax mapping rule (van Valin, 2005); Section 5 explains the adjunct island effects as well as parasitic gap effects within this new analysis and suggests future extensions.

2 Coherence and Conjuncts

In this section I summarize previous contributions to the coherence-based theory. As Ross (1967) first notes, normally extraction from conjuncts occurs across the board, which means the extracted element needs to correspond to a position in both conjuncts. This is exemplified in (1). [1]

(1) What book$_i$ did you buy t$_i$ and read t$_i$?

Aside from such symmetric extraction, nevertheless, asymmetric extraction can also take place across conjuncts, such as the following (Kehler, 2002):

(2) a. That's the stuff$_i$ that the guys in the Caucasus drink t$_i$ and live to be a hundred.
b. How many lakes$_i$ can we destroy t$_i$ and not arouse public antipathy?
c. Which liquor$_i$ did you go to the store and buy t$_i$?

(2a) expresses a causal relation between the two conjuncts. (2b) expresses a violated expectation (concessive) relation between the two conjuncts: the second event is an unexpected development given the first event. In (2c) the first conjunct provides a scene (frame, setting, etc.) for the second event to occur. In all these cases, the element being extracted corresponds to the position within one conjunct, but not the other. Symmetric extraction is always possible in these relations, but (2) indicates that the constraint for extraction to be across the board is not categorical.

According to Kehler (2002, 2008), which draws upon previous approaches, symmetric and asymmetric extraction from conjuncts can receive a unified coherence-based explanation. This is based on Hobbs' (1979) argument that one of the three coherence relations (Parallel, Cause-Effect and Occasion) must be inferred and established between conjuncts for them to be asserted felicitously. Establishing a coherence relation requires that a link be identified between the propositions denoted by the utterances in a passage. According to Hobbs, the establishment of Parallel, Cause-Effect and Occasion relations proceeds as follows, respectively:

(3) Parallel: Infer P(a_1, a_2,...) from the assertion of S_1 and P(b_1, b_2,...) from the assertion of S_2, where for some property vector q_i, $q_i(a_i)$ and $q_i(b_i)$ for all i.
(4) Explanation: Infer P from the assertion of S1 and Q from the assertion of S2, where normally $Q \rightarrow P$.
(5) Occasion: Infer a change of state for a system of entities from the assertion of S_2, establishing the initial state for this system from the final state of the assertion of S_1.

To explain extraction behaviors, Kehler further proposes that only potential common topics for the conjuncts may undergo extraction. A potential common topic is identified when a given coherence relation is established.

[1] constructions such as (1) will be henceforth referred to as ATB

Take the Parallel relation for example. Once we infer (3) and establish a Parallel coherence relation between propositions S_1 and S_2, it follows that we can determine candidates for a common topic in a programmatic way. First, any assertion of a proposition in a particular situation can be construed as being about a particular entity. Assume a situation where the assertion of S_1 (*i.e.* denoting a relation P (a_1, a_2, ..)) is construed as being about a_1, *i.e.* S_1 is a statement that a1 bears the property of in the relation P with other entities. Then we infer that parallel to S_1, the assertion of S_2 under the same situation is a statement that b_1 is related by the relation P to other entities. Here a_1 and b_1 occupy the same formal position within the n-place predicate P, and they share the property of q_1. Kehler argues that we can then think of the common topic here as the superordinate category of both a_1 and b_1. If a_1 and b_1 are totally identical, there will be no another entity that serves as a superordinate category, so the common topic will be a_1 (=b_1). This is exactly the case of across-the-board extraction. Moreover, a_1 and b_1 simply needs to share the property q1 and may be specified differently in other properties. Then, the common topic can be an entity that is specified with the property q1 but is unspecified for the other properties. The following example is given by Kehler to illustrate this scenario:

(6) Speaking of reading materials, John bought the books and Bill bought the magazines.

Here the preposed element reading materials does not correspond to a slot within either of the conjuncts, giving rise to a "gapless" extraction. It is more general than the books and the magazines, hence subsuming both, who may only share partial identity with each other (Lakoff, 1986).

Cause-Effect Relations determine potential common topics in a similar manner. As (4) shows, the second event of Q forms a causal link with the first event of P. Thus, the final state of Q will connect with the P-event and serves also as P's result state. Given a situation where the assertion of P is construed as about an entity within P, Q will be part of the coherent scenario that centers around that entity, even if the entity doesn't correspond to any constituent in the representation of Q. As a result, this entity qualifies as a common topic in this scenario, so that it can be extracted from its conjunct. The inference process for Violated Expectation is similar, except that we infer that P leads to Q although normally P does not cause Q. In this sense, concessive relation can be seen as a negation of causal relation (König and Siemund, 2001).

The inference process underlying Occasion relation is different. In this relation, the event denoted by the first conjunct provides the scene/circumstance for the event denoted by the second conjunct. As the two events are temporally contiguous, this scene-setting is construed as a transition where the initial state of the second event connects with the final state of the first event. As a result of this transition, as we reach the end state of the first event, we infer that a change of state takes place so that we are now in the initial state of the salient event. For each circumstance, the second event is uniquely identified by a particular connection that differs it from similar events of the same type. Thus, the

two connected events can express a coherent scenario centered around a salient entity in the second event. On the other hand, as the connection is seen as the first event preparing the scene to evoke the salience of the second event, not the other way round, the internal entities within the first event are not seen as salient within this coherent scenario.

In general, both symmetric and asymmetric extraction from conjuncts are sensitive to a coherence condition: an entity is extractable if it is salient throughout both conjuncts, which can be achieved when the connection of two conjuncts constitute a coherent scenario about that entity.

3 Chinese Adjunct Data

Previous literature often assumes that overt extraction from adjuncts obey the syntactic island conditions (Lin, 2005; Ting and Huang, 2008). This is motivated by data such as the following: both (8) and (9) are relative constructions where the relative head corresponds to a constituent within the adjunct part of the relative clause. (8) is unacceptable, whereas judgment in (9) is good.[2]

(8) *[Laoban [mianshi-wan ti yihou] chuqu chi fan-le] de yingpinzhe$_i$
Boss interview-ASP after go.out eat meal-ASP REL applicant
'The applicant$_i$ [that the boss went out for meal [after (he) interviewed _$_i$]]'

(9) [Laoban [mianshi-wan ti yihou] jueding luyong ti] de yingpinzhe$_i$
Boss interview-ASP after decide recruit REL applicant
'The applicant$_i$ [that the boss decided to recruit _textsubscripti [after (he) interviewed _$_i$]]'

Syntactic approaches argue that this contrast is due to that in (9), a second gap occurs in the matrix part of the relative clause, creating a parasitic gap environment. Furthermore, even in a parasitic environment, if the matrix clause is further embedded in a relative clause, or if the adjunct clause is recursively embedded in another adjunct clause, judgment will be downgraded again, echoing the cases in English (Contreras, 1984; Xu, 1990; Cinque, 1990). Although I agree with the judgment patterns as mentioned above, I argue that this is not yet the full picture. Instead, I argue that the following sentences in (10-11) are also acceptable (the native speakers I consulted with unanimously agreed with my judgment), posing a severe challenge to the syntactic approach.[3]

(10) [Zhangsan [yujian ti yihou] jueding gen yuanlai nüyou fenshou] de nei-ge nvhair$_i$
Zhangsan meet after decide with then girlfriend break.up REL DEM-CLF girl
'The girl$_i$ [that Zhangsan decided to break up with his then girlfriend [after (he) met _$_i$]]'

(11) [[Chubanshang jueding chuban ti yihou] zuozhe fan'er youyuqilai] de nei-ben xiaoshuo$_i$
Publisher decide publish after author nevertheless be.reluctant REL DEM-CLF novel

[2] ASP:aspectual marker; REL: relativizer
[3] DEM:demonstrative; CLF: classifier

'The novel$_i$ [that the author became reluctant [after the publisher decided to publish _$_i$]]'

In (10), the event denoted by the adjunct and the event denoted by the matrix clause are not only temporally contiguous, but also form a causal link. In (11), the event denoted by the matrix clause occurs contrary to the normal expectation of the adjunct event. Thus these examples resemble the asymmetric extraction behaviors in English conjuncts in Cause-effect and Violated Expectation relations. Furthermore, in the following examples that I construct, if we force a reading where the temporal adjunct and the matrix stand in the Hobbsian Parallel relation, and disallows a Cause-Effect reading (i.e., watching movie follows as a consequence of me reading the book), then a symmetric extraction requirement becomes operative: judgment is downgraded both when extraction is from the adjunct and when it is from the matrix.

(12) a. *Nei-ben shu$_i$, wo du-le ti yihou kan-le yi-bu dianying.
DEM-CLF book, I read-ASP after watch-ASP one-CLF movie
'That book$_i$, I watched a movie after (I) read _$_i$.'

b. *Nei-bu dianying$_i$, wo du-le yi-ben shu yihou kan-le ti.
DEM-CLF movie$_i$, I read-ASP one-CLF book after watch-ASP
'That movie$_i$, I read a book after (I) watched _$_i$.'

Another similarity concerns gapless extraction. As has mentioned above, in extraction from conjuncts the common topic may be less specific than the pair of entities in both conjuncts and simply be a superordinate category for the pair, allowing gapless extraction. As the following illustrates, a similar possibility for the extracted element to be a superordinate category also obtains in extraction from adjuncts, giving rise to gapless constructions like (13a-b).

(13) a. Shuiguo, Zhangsan xihuan xiangjiao erqie Lisi xihuan pingguo.
Fruit, Zhangsan enjoy banana and Lisi enjoy apple
'Fruit, Zhangsan enjoys banana and Lisi enjoys apple.'

b. Fabiao, Zhangsan ji xie-le lunwen you chu-le shu.
Publications, Zhangsan both write-ASP paper and publish-ASP books
'(Speaking of) publications, Zhangsan has both written papers and books already.'

4 Interclausal Hierarchy and Clause Linkage

The new data in (10-13) show that extraction from adjuncts and from conjuncts need to be characterized in a uniform way: here a common set of semantic relations are encoded by subordinative structures in one language and by coordinative structures in another language, but the extraction possibilities are the same for the two structures. This suggests that the extraction condition is sensitive to the semantic relations, not to syntactic structures.

Therefore, at least for temporal adjuncts in Chinese, we can rephrase the adjunct island constraint as a semantic filter, which dictates that given a coherence

relation, if the extracted element does not qualify as the common topic under a coherent scenario, the sentence will be infelicitous.

Before proceeding, we need to consider a reanalysis approach: Maybe the adjunct clauses in these examples are actually coordinative structures. Accept coherence for conjunct structure, but still maintain adjunct island conditions. This notion has been proposed for English (Huybregts and van Riemsdijk, 1987; Williams, 1990) to capture the observations that parasitic gap extraction from across adjuncts share strong semantic similarities with across the board extraction from across conjuncts (Ross, 1967; Grosu, 1980; Torris, 1983).

Williams suggests that the more a subordinative structure exhibit syntactic and semantic symmetry, the closer it would be construed as coordinative and the higher acceptability its PG will receive. Thus the acceptability hierarchy for different parasitic gaps (henceforth: PG) is reduced to a gradable scale of "coordinativeness". The PG-ATB parallelism, thus, is accounted for because a subset of subordinative structures are actually subject to the licensing rules governing ATB.

This sort of coercion is undesirable. First, in short of a precise characterization, it is never clear how to coerce a coordinative structure from a subordinative structure, nor is it clear what it means to talk about "graded" coordinativeness. In particular, this coercion idea is originally motivated for the ATB-PG similarities, but since coordinative structure allows asymmetric extraction, we would have to coerce subordinative structure that allows asymmetric extraction to be coordinative, too. It thus seems that most subordinative structures share some "coordinativeness" to a certain degree. And since the island-inducing adjunct examples can also be construed as coordinative structure that bears an Occasion coherence relation, there might be no uncontroversially subordinative structures upon which adjunct island conditions apply.

One further piece of evidence is that a reanalysis approach is ill fit to account for the fact that the use of coordination to encode coherence relations in Chinese is quite constrained. Chinese encodes coordinative linkage by the sentential conjunction marker erqie, which corresponds to the sentential use of and in English. We see that English and is able to encode all the coherence relations in (1). In Chinese, as (14a-d) shows, all four coherence relations can be encoded by subordinative linkage, but only Parallel (13a) and Violated Expectation (14b) relations can also be optionally expressed by coordinative linkage. This option is not available for the Cause-Effect (14c) and Contiguity Relations (14d), which obligatorily employ subordinative linkage strategy.

(14) a. Neiben shu_i, ni mai-le ti yihou/erqie du-guo?
DEM-CLF book you buy-ASP after/CONJ read-EXP
'Which book_i, did you read $\_\_i$ after (we) bought $\_\_i$?'

b. Duoshao-tiao hupo_i, women keyi huidiao ti yihou /erqie bu yinqi gong fen?
How.many-CLF lake we can destroy after/CONJ NEG arise public outcry
'How many lakes_i, can we not arise public outcry after (we) destroy $\_\_i$?'

c. Nei-ge jiushi [gaojiasuo de ren he-le ti yihou/*erqie huo dao yibaisui] de dongxi_i.

DEM-CLF BE Caucasus REL person drink-ASP after/CONJ live till one.hundred.age REL stuff
'That is the stuff$_i$ that people in Caucasus lived till a hundred years old after (they) drank _$_i$.'

d. Zhei-ge weishiji$_i$, wo qu shangdian yihou/*erqie mai-de ti.
DEM-CLF whiskey I go store after/CONJ buy-ASP
'This whiskey$_i$, I bought _$_i$ after (I) went to the store.'

If, as a renalaysis approach would argue, the temporal adjunct and the matrix clause linked by yihou 'after' is actually a coordinative structure, why would this coerced coordinative structure more liberal in allowing extraction than a canonical, true coordinative structure?

On the other hand, I argue that if we are to accept a semantic extraction condition, this restraint in Chinese coordinative structure can be captured by an independently motivated semantics-syntax mapping hierarchy that addresses the crosslinguistic differences in clause linkage strategies.

Typological literature has formulated a set of coherence relations that denote the connection between propositional units crosslinguistically (Lehmann, 1988), which are compatible with Hobbs' formulations. Specifically, the semantic relations of causation, concession and circumstance used in the typological tradition match with result, violated expectation and occasion respectively (Grote et al, 1995). The typological terms tend to focus on the fundamentally semantic nature of these relations (König and Siemund, 2000), whereas the coherence terms underscore their cognitive and rhetorical/interactional nature (Couper-Kuhlen and Thompson, 2000). Leaving aside this distinction, it is noteworthy that different languages employ different clause linkage strategies to encode these semantic relations in their syntax (Couper-Kuhlen and Kortmann, 2000). The choice is subject to language-internal conventionalization, but also subject to a general hierarchy (Silverstein, 1976; Givon, 1980; Foley and van Valin, 1984; van Valin and LaPolla, 1997; van Valin, 2005), such that more cohesive semantic relations between propositional units tend to be expressed by stronger linkage strategies.

The cohesiveness of an inter-propositional semantic relation can be understood as how connected are the two propositional events relative to each other. For example, a Parallel coherence relation normally comprises of two temporally simultaneous or sequential states of affairs, which are expressed by two discrete events/actions. However, in an Occasion relation, the scene-setting event is interpreted as a spatial/temporal parameter of the primary event. In this sense it expresses one facet of a single action. Similarly, in a Result relation, the result event is construed as expressing the end state of a causal chain, thus also expressing a facet of a single action (van Valin, 2005: 208). In general, measured by the degree to which the propositional units depict facets of a single action/event van Valin (2005) argues for a general cohesiveness scale. Limiting to the four coherence relations discussed in this paper, we can assume the following ranking of cohesiveness:

(15) Occasion > Result > Violated expectation > Parallel Where > denotes 'more cohesive than'

On another end, the strength of clause linkage strategies pertains to the way that juncts (i.e. sentences, clauses, phrases, etc. see van Valin, 2005: 209) are linked together. A set of crosslinguistically robust criteria (properties) have generally set apart subordination and coordination as two distinct clause linkage structures (Lehmann, 1988; Couper-Kuhlen and Kortmann, 2000). Although closer statistical studies reveal that the set of properties that set the two structures apart are not clear-cut, as a tendency there is still a type of "subordinative" prototype which can be distinguished from coordination type, differing crucially in the possibilities of illocutionary scope marking, tense scope marking, flexible position, etc. (Bickel, 2010) Importantly, we can take these behaviors as evidence to indicate that subordinate structure as a type tends to be formally more integrative than coordinative structures, represented as follows.

(16) subordination $\succ$ coordination. Where $\succ$ denotes 'stronger than'.

Importantly, this semantics-syntax mapping hierarchy cannot be one-on-one correspondence, as the same semantic relation may well be mapped onto two different linking strategies for two languages, and even within one language, two linking strategies can be available for the same semantic relations.

Rather, for any given language, this hierarchy should be understood as implicational:

(17) If $R_1 > R_1$, then $S_1 \succeq S_2$.
For two semantic relations R_1 and R_1, and two clause linkage strategies S_1 and S_2 such that R_1 maps to S_1 and R_2 maps to S_2. $>$ denotes 'more cohesive than', $\succeq$ denotes 'at least as strong as'

Independent survey of typologically distinct languages also supports the validity of this implicational hierarchy crosslinguistically (Kockelman, 2003). The Chinese data in (14) now can be readily accounted for by this implicational hierarchy. First, because the less cohesive violated expectation and parallel relations can be encoded by subordination in Chinese, the hierarchy correctly predicts that the more cohesive Occasion and Result relations are also encoded by subordination. Second, while violated expectation/parallel can be encoded by coordination, occasion/result cannot. This is also predicted since the implicational hierarchy totally allows that stronger semantic relations do not map to a weaker clause linkage strategy.

5 Conclusion

We now possess the apparatus to explain the extraction from adjuncts in Chinese in a uniform way.

First, for extraction that involves gaps, both the adjunct and the matrix clause must contain a gap when the two events stand in a parallel relation for a coherence scenario to be established for the gapped constituent, thus correctly predicting the phenomena in (12a-b), where extraction from only one of the sentences (adjunct or matrix) is always bad. When the two events stand in a

cause-effect relation as in (10-11), they constitute a connected scenario that centers around an entity within the first event, making asymmetric extraction from adjuncts possible. When the two events are related in an Occasion relation as in (8), adjunct island effects are predicted to arise, due to the fact that the first event provides the scene for the second event, so that only entities in the second event are salient in the coherent scenario.

As the condition on extraction is defined in terms of coherent scenarios, this theory predicts that parasitic gap extraction is always good: in a parasitic gap construction, the two events share one common event participant (i.e. the entity denoted by the extracted element), which means we know that a common entity remains salient for both events, thus always satisfying the requirement for the two events to constitute a coherent scenario. For example, in (9), repeated below: The relative head, the applicant, belongs to both adjunct-denoted interviewing event and the matrix-denoted recruiting event. Here the two events stand in one of the coherence relations, i.e. They constitute an Occasion scenario. Thus, the applicant is not only salient within the scene-setting event but also remains salient within the primary event, so that the coherent scenario as a whole centers around it.

(18) [Laoban [mianshi-wan ti yihou] jueding luyong ti] de yingpinzhe$_i$
Boss interview-ASP after decide recruit REL applicant
'The applicant$_i$ [that the boss decided to recruit _$_i$ [after (he) interviewed _$_i$]]'

Importantly, this situation is only part of the broader observation that in all coherence relations, including both those that allow for symmetric and asymmetric relations, it is always the case that PGs are good. Thus, there is no need to stipulate a specialized constraint for PG, as opposed to other adjunct constructions. Rather, the circumvention effect of PG follows naturally from the coherence conditions.

Finally, for gapless cases where an element corresponds to none of the clauses, the same mechanisms apply as in cases with gapped extraction in all the above relations. A case of Parallel relation has been mentioned already in (13a), repeated below.

(19) Shuiguo, Zhangsan xihuan xiangjiao erqie Lisi xihuan pingguo.
Fruit, Zhangsan enjoy banana and Lisi enjoy apple
'Fruit, Zhangsan enjoys banana and Lisi enjoys apple.'

Here *fruit*, as a superordinate category of both *banana* and *apple*, is salient in both events. Thus, the coherent scenario centers around *fruit*. Similarly, in other coherence relations, e.g. in a Result relation, when an entity is the superordinate category of another entity that appears within the first event (cause event), the whole scenario may also center around the superordinate entity, as (20) illustrates.

(20) Chubanshang jian-guo zuozhe yihou like gaibian-le taidu de nei-ben xiaoshuo
Publisher meet-ASP author after immediately change-ASP attitude REL DEM-CLF novel

'The novel that the publisher changed their attitude right away after they met with the author'

As my theory is formulated in general semantic terms, it is my expectation that it extends to all languages in a predictive manner, although I do not attempt to make such an extension in the face of the complexities with how different languages encode subordination differently. It is worth noting, though, that the adjunct examples Truswell (2007, 2010) gives may be compatible with a coherence explanation: Truswell discusses two scenarios where the adjunct denotes a preceding event that causes or describes the matrix event. As such the matrix event can be construed as a coherent part of the scenario centering around a salient entity within the adjunct event, denoting an end state or a result state. The fact that Truswell's examples involve non-bare adjuncts that are tenseless and nonfinite (hence no Chinese counterparts can be found), but might receive a similar explanation, is a particularly encouraging sign.

Acknowledgments. I am grateful to the three anonymous reviewers for the ESSLLI student session for their useful suggestions and encouragement. I also appreciate the audience at the 2014 LSA annual meeting and the Linguistics Colloquium at the Linguistics Department, SUNY Buffalo for their feedback. I am particularly indebted to Jun Chen and Lihua Xu for their overall support and helping me with the data, as well as to a group of SUNY Buffalo undergraduate students (Mandarin native speakers) for supplying their judgments. The usual disclaimer applies.

References

1. Beard, Robert. Decompositional composition. Natural Language and Linguistic Theory 9(2), 195–229. (1991)
2. Cinque, Gullielmo. Types of A-bar dependencies. Cambridge. (1990)
3. Contreras, Heles. A Note on Parasitic Gaps. Linguistic Inquiry 698-701. (1984)
4. Couper-Kuhlen, Elisabeth., Thompson, Sandra. Concessive Patterns in conversation. Walter de Gruyter. (2000)
5. Deane, Paul. Limits to attention: A cognitive theory of island phenomena. Cognitive Linguistics2(1), 1-63. (1990)
6. Foley, William, van Valin, Robert. Functional Syntax and Universal Grammar. Cambridge. (1984)
7. Givon, Talmy. The binding hierarchy and the typology of complements. Studies in Language Groningen 4(3), 333-377. (1980)
8. Goldsmith, John. A Principled Exception to the Coordinate Structure Constraint. In: Papers from the 21st regional meeting of the Chicago Linguistics Society. pp:133-143. (1985)
9. Grosu, Alexander. Should there be a (restricted) rule of conjunction reduction? Linguistic Inquiry12(1),149-150.
10. Heestand, Dustin, Xiang, Ming, Polinsky, Maria. Resumption Still Does Not Rescue Islands. Linguistic Inquiry 42(1), 138-152. (2011)

11. Hofmeister, Philip., Casasanto, Lisa., Sag, Ivan. Islands in the grammar? Standards of evidence (2014)
12. Hu, Jianhua., Pan, Haihua. Decomposing the aboutness condition for Chinese topic constructions. Linguistic Review26, 371-384. (2009)
13. Huang et al. Syntax of Chinese. Cambridge. (2009)
14. Huybredgts, Reny., van Riemsdijk, Henk.: Parasitic Gaps and ATB. In:Proceedings of NELS. 181–184. (1986)
15. Kehler, Andrew. Coherence, reference, and the Theory of Grammar. CSLI. (2002)
16. Kehler, Andrew. Coherence and Coreference Revisited. Journal of Semantics25(1), 1-44.
17. Kockelman The Interclausal Relations Hierarchy in Q'eqchi' Maya. International Journal of American Linguistics69(1), 25-48. (2003)
18. Koenig, Ekkehard.,Siemund, Peter.Causal and concessive clauses: Formal and semantic relations. in Topics in English Linguistics33. pp: 341-360. (2000)
19. Kuno, Susumu. The position of relative clauses and conjunction. Linguistic Inquiry 117-136. (1974)
20. Lakoff, Robin. Frame Semantic Control of the Coordinate Structure Constraint. Berkeley. (1982)
21. Lehmann, Christian. Towards a Typology of Clause Linkage. In: Clause Combining In Grammar and Discourse: 181-225. (1988)
22. Lin, Jonah. Does wh-in-situ license parasitic gaps? Linguistic Inquiry 36:298-302. (2002)
23. Postal, Paul.Parasitic gap and the across-the-board phonemena Linguistic Inquiry 735-754. (1993)
24. Pustejovsky, James.The Generative Lexicon. MIT Press. (1993)
25. Schmerling, Susan. Asymmetric conjunction and rules of conversation. Syntax and Semantics3: 211-231. (1972)
26. Silverstein, Michael. Hierarchy of features and ergativity. In: Grammatocal categories in Australian languages. pp: 112-171. (1976)
27. Ting, Jen., Huang, Yuchi. Some Remarks On Parasitic Gaps in Chinese. Concentric34, 127-152. (2008)
28. Truswell,Robert. Extraction from Adjuncts and The Structure of Events. Lingua117(8), 1355-1377. (2007)
29. Truswell,Robert. Events, Phrases and Questions. Oxford. (2010)
30. van Valin, Robert. Exploring the Syntax-Semantics Interface. Oxford.(2005)
31. van Valin, Robert., and LaPolla,Randy. Syntax: Form, Meaning and Function. Cambridge. (1997)
32. Williams, Edwin. The ATB theory of parasitic gaps. Linguistic Review 6(2): 265-279. (1990)
33. Liejiong Xu. Are They Parasitic Gaps. In: Grammar in Progress: Glow essays for Henk van Riemsdijk. pp:455-461. (1990)

Achievement Predicates and the Slovenian Imperfective Aspect*

Maša Močnik

ILLC, MoL, Universiteit van Amsterdam,
P.O. Box 94242, 1090 GE Amsterdam, The Netherlands
masa.mocnik@gmail.com

1 Introduction

Achievements are characterised as telic, instantaneous events, illustrated by sentences such as *John arrived* or *John reached the top* [13]. Altshuler [2] studies the Russian imperfective aspect and the English progressive aspect, and shows that the two pattern very differently with respect to achievements: Russian imperfective achievements come with a culmination entailment, as in (1), whereas the corresponding English progressive achievements do not.

(1) K nam priezžal otec domoj, (# no on ne smoh najti naš dom)
to us arrive$_{IPF.PST}$ father home but he not able find our house
'Father came to see us at home, but was unable to find our house.' (after [2, p. 46, ex. (13)])

(2) Father was coming to see us at home, but was unable to find our house. [2, p. 46, ex. (14)]

Though the Slovenian imperfective aspect shares many similarities with its Russian relative, see [5] and §2, it patterns like the English progressive aspect with achievement situation types:[1]

(3) Janez je dosegal vrh, a ga ni dosegel.
John aux reached$_{IPF}$ top but him neg-aux reached$_{PF}$
'John was reaching the top, but didn't reach it.'

The goal of the present paper is to explore the use of the Slovenian imperfective aspect and propose a semantics for it that can account for this type of data. The relevant properties of the Slovenian imperfective are presented in §2: the well-known Slavic *konstatacija fakta* in §2.1 and imperfective performatives in §2.2. Section §3 presents Altshuler's analysis of the Russian imperfective and the English progressive. I show in §4 that neither proposal can be straightforwardly

* Thanks to Frank Veltman, Daniel Altshuler, Benjamin Sparkes, the Szklarska Poreba Workshop 2014 audience, and anonymous reviewers for comments and discussion.

[1] It is unfortunately not possible to construct a minimal pair with Altshuler's (1) and (2) since there is no imperfective of *prispeti* (*arrive*$_{PF}$) with the same meaning.

adopted for Slovenian, and propose a constraint on the semantics of the Russian imperfective that can account for it. I conclude the paper in §5 by relating the proposed semantics to a recent theory of performativity proposed in [4].

2 Slovenian Imperfective Aspect

The Slovenian tense and aspect system maintains the distinction between perfectivity and imperfectivity in its three tenses – the simple present and the periphrastic past and future.

(4)	a.	pisal je	piše	pisal bo	
		write$_{PTC.IPF}$ is	writes$_{IPF}$	write$_{PTC.IPF}$ is$_{FUT}$	
		'he was writing'	'he is writing'	'he will be writing'	
	b.	na-pisal je	na-piše	na-pisal bo	
		PF-write$_{PTC.IPF}$ is	PF-writes$_{IPF}$	PF-write$_{PTC.IPF}$ is$_{FUT}$	
		'he wrote'	'he writes'	'he will write'	

Slovenian verbs are traditionally considered inherently perfective or imperfective [14],[2] though they can change their aspectual class via imperfective suffixation or perfective prefixation. As is well-known in Slavic, grammatical and lexical aspect are closely interrelated and the use of a perfectivising affix can result in a shift of aspectual class. (I gloss *na-* above as a purely perfective prefix for simplicity's sake.) Telicity is a property of Vs in Slavic [7], so we can speak about, for instance, 'achievement verbs' or 'achievement predicates'.

Though the Slovenian imperfective aspect has a variety of uses,[3] the following sections focus on two. I first discuss the so-called *konstatacija fakta*, which expands on the data presented in the introducton, and then turn to performative utterances, as they provide an important clue (see §4) to understanding the semantics of the Slovenian imperfective. For a good overview of some of the other uses, see [6].

2.1 Konstatacija fakta

The imperfective aspect can be used to signal a 'single, completed action' ([8], cited in [5]). Dickey, who refers to this as *the imperfective general-factual*, characterises it as 'the use of an [imperfective] past-tense verb form simply to confirm the occurrence of an action, without reference to specific circumstances' [5, p.

[2] There are also the so-called biaspectual verbs like *telefonirati* (to telephone), cf. [14].

[3] See [5] for habituality, iterativity, and the historical present. The Slovenian imperfective also appears with all three temporal relations, viz. simultaneity, anteriority and, sometimes, posteriority ('action in sequence' in [5]).

Like Altshuler [2], I must leave states aside for future research (!) – it remains to be seen how the stage-based analyses of the Slovenian imperfective (here) and the Russian imperfective (in [2]) can be extended to states, which are argued not to have stages (see [11], for instance).

95]. His investigation leads him to the conclusion that the general-factual use with accomplishments appears in all Slavic (including Slovenian). The utterances in (5), for instance, imply the culmination of the accomplishment event: it is implied that the speaker has finished the book and that the tree has been decorated. Since the inference is cancellable, e.g. with *but I never finished it* in (5a), this is indeed an implicature.

(5) a. Kot najstnik sem bral *Hlapce*.
as teenager aux read$_{IPF}$ Serfs
'As a teenager, I read *Serfs*.' (after [5])

b. Božično drevo smo okraševali (že) prejšnji teden.
Christmas tree aux decorated$_{IPF}$ (already) previous week
'We decorated the Christmas tree (already) last week.'

Dickey also concludes that the imperfective general-factual with achievements does not appear in all Slavic languages – it is restricted to Russian, Belarusian, Bulgarian and Ukrainian.[4] Consider his examples for Slovenian:

(6) a. Kot otrok sem padel / #padal s tega drevesa.
as child aux fallen$_{PF}$ fallen$_{IPF}$ from this tree
'As a child, I fell from this tree.'

b. Enkrat je dobil / #dobival ukor zaradi zamude.
once aux gotten$_{PF}$ gotten$_{IPF}$ reprimand because-of lateness
'He once got a reprimand for being late.'

c. Ali si se že kdaj spotaknil / #spotikal na ulici?
Q aux refl already ever trip$_{PF}$ trip$_{IPF}$ on street
'Have you ever tripped on the street?

(based on [5, p. 101])

In contrast to Russian (1), the inference that there was a single, completed event can only be obtained with the perfective, while the imperfective form expresses iteration/habituality (e.g. *As a child, I kept falling / used to fall from this tree*) or focuses on the preparatory stage (e.g. *As a child, I was falling from this tree when I had a heart attack*). The latter effect is also discussed in [10] for the following example:

(7) Janez je dosegal vrh. [10, p. 96]
John aux reached$_{IPF}$ top
'John was reaching the top.'

As its English translation, the sentence focuses on the interval before the culmination: John was on his way to reaching the top [10, p. 96]. There is neither entailment nor implicature that John has reached it. Note also that both (5a) and (7) point towards the fact that the Slovenian imperfective need not denote a whole event, since cancellation would otherwise not be possible.

[4] Polish is borderline, cf. [5, p. 124].

2.2 Imperfective Performative Utterances

English performative utterances typically come in the simple present – their progressive counterparts lack the performative inference. Compare:

(8) a. I promise to come tomorrow.
b. ?? I'm promising to come tomorrow.

The speaker performs the action named by the verb by uttering (8a), i.e. he makes the promise to come tomorrow. The sentence in (8b), on the other hand, is reported to be odd. It is indeed '[a] well-known generalization [...] that utterances in the progressive cannot (usually) be used performatively' [4, p. 162]. Slovenian is not in immediate contradiction with this claim (though see §5) since the claim concerns the progressive aspect, often considered a subcategory of the imperfective aspect (cf. [3, p. 25]).[5] Consider now the situation in Slovenian:

(9) a. Obljubim, da bom prišel jutri.
promise$_{1SG.PRES.PF}$ that aux$_{FUT}$ come$_{PTC.PF}$ tomorrow
'I promise that I will come tomorrow.'
b. Obljubljam, da bom prišel jutri.
promise$_{1SG.PRES.IPF}$ that aux$_{FUT}$ come$_{PTC.PF}$ tomorrow
'I promise$_{IPF}$ that I will come tomorrow.'

Both (8a) and (8b) are performative utterances – the speaker makes a promise to come the following day. Similarly with other aspectual pairs, e.g. *zahvaliti*$_{PF}$ *se / zahvaljevati*$_{IPF}$ *se* (to thank), *prisečiPF / prisegati*$_{IPF}$ (to swear) or *priznati*$_{PF}$ */ priznavati*$_{IPF}$ (to confess). The difference between the perfective and the imperfective is usually a matter of register – imperfective performatives are typically used in more formal settings, cf. [15]. Nevertheless, the fact that the Slovenian imperfective is naturally used in performative utterances shows that it is compatible with the inference of a whole, complete event, such as the promise to come tomorrow, see §4.

3 Setting Up the Stage

Altshuler's [2, pp. 47–50] analysis of the Russian imperfective and the English progressive builds heavily on Landman's [9] intensional, event-based analysis of the English progressive. Landman assumes that sets of events are ordered by two relations: *part-of* and *stage-of*. The part-of relation is a broader notion: listening to the radio can be part of making pancakes, but it is not (normally) one of its stages [11, p. 47]. A stage is understood to be a 'less developed version' of an

[5] Alternatively, the claim could be interpreted as concerning the English progressive only. The reader familiar with the proposal in [4] will have noticed, however, that any progressive perfomative poses problems for the theory. I return to this issue in §5 with Slovenian.

event [9, p. 23], such as mixing pancake batter [11, p. 47]. I use the symbol $\leq$ for the *part-of* relation and $\sqsubseteq$ for the *stage-of* relation.

The *stage-of* relation between two events is encoded in [2] with the help of the STAGE operator (STAGE for English and STAGE* for Russian). In a slightly changed formalisation, Altshuler's idea is the following:

(10) $[\![IPF_{RU}(P)(e')(w^*)]\!]^{\mathcal{M},g} = 1$ iff $[\![\exists e \exists w\ STAGE^*(e', e, w^*, w, P)]\!]^{\mathcal{M},g} = 1$ iff there is some g' that differs from g in at most the values for e and w such that:

a. the history of $g'(w)$ and $g'(w^*)$ is the same up to and including $\tau(g'(e'))$

b. $g'(w)$ is a reasonable option for $g'(e')$ in $g'(w^*)$

c. $[\![P(e)(w)]\!]^{\mathcal{M},g'} = 1$

d. $g'(e') \leq g'(e)$ (modified from [2, p. 49])

The Russian imperfective operator, like the English progressive, is assumed to be a partitive operator, i.e. a relation between properties of events and their parts (with respect to worlds). Given a model $\mathcal{M}$ and an assignment function g, when combined with a property of events P, an event e' and a world of evaluation w^*, the Russian imperfective requires, roughly speaking, there to be an event e and a world w such that e' in w^* is a 'stage' of e in w. More precisely, there must be an assignment function g' that differs from g in at most the values for e and w such that: (a) the world denoted by w and the world denoted by w^* have the same history until the end of the runtime of the event denoted by e' (the temporal function τ maps events onto their runtimes), (b) the world denoted by w is a reasonable option for the event denoted by e' in the world denoted by w^*, i.e. 'there is a reasonable chance on the basis of what is internal to [the event denoted by e'] in [the world denoted by w^*] that [the event denoted by e'] continues in [the world denoted by w^*] as far as it does in [the world denoted by w]' [9, p. 25], (c) the event denoted by e has the property P in the world denoted by w (e.g. it is a making-of-pancake event), and (d) the event denoted by e' is a non-proper part of the event denoted by e.[6]

The semantics of the English progressive is almost identical to (10):

(11) $[\![PROG_{EN}(P)(e')(w^*)]\!]^{\mathcal{M},g} = 1$ iff $[\![\exists e \exists w\ STAGE(e', e, w^*, w, P)]\!]^{\mathcal{M},g} = 1$ iff there is some g' that differs from g in at most the values for e and w such that: (a)–(c) are as in (10), and (d) $g'(e') < g'(e)$. (modified from [2, p. 49])

The difference lies in clause (d): the event denoted by e' must be a proper part of the event denoted by e. This distinction enables one to explain data such as (1) and (2). Recall that Russian imperfective achievements, e.g. (1), come

[6] For simplicity's sake, as noted in [2], IPF_{RU} does not relate an event to a topical time, as in e.g. [1], and STAGE* does not include Landman's [9] continuation branch function.

with a culmination entailment, whereas English progressive achievements, e.g. (2), do not. Altshuler proposes that achievements are single, atomic stages with no proper stages. He thinks of an atomic stage as a stage that 'develops into itself in the world of evaluation (and presumably every other possible world)' [2, p. 49]. This assumption explains why Russian imperfective achievements entail event culmination: the atomic stage *is* the event itself. Note that clause (10d) is then trivially fulfilled. On the other hand, achievements not having proper stages leads to a clash in clause (11d). Altshuler appeals to coercion as a repair strategy: English achievements are restructured into accomplishments (along the lines of [11]) and the event denoted by e' becomes a proper part of the new accomplishment.[7] In the following section I return to this in more detail.

4 Accounting for the Slovenian Imperfective

The Slovenian imperfective operator (henceforth IPF_{SLO})[8] is a partitive operator, as it can denote incomplete events. This was informally observed in §2.1, where it was possible to cancel event culmination in examples (5a) and (7). The latter suggests that IPF_{SLO} can encode the strict part-of relation ($<$) in clause (d). I propose, however, that IPF_{SLO}, like its Russian relative, encodes the non-strict relation ($\leq$) due to it naturally appearing in performative utterances, see §2.2. If IPF_{SLO} is *compatible* with denoting a whole event (via $\leq$), we are in a better position to potentially explain why an imperfective utterance can have the performative inference, say, that a promise has been made, see §5.

The lexical entry obtained at this point is identical to (10). We can therefore straightforwardly explain the *konstatacija fakta* use of IPF_{SLO} with accomplishments. Clause (10d) states that the event denoted by e' represents *at least* some of event e, possibly it is the whole event. As for Russian in [2], we can argue that accomplishments have at least two stages, which is why what is entailed is the culmination of a smaller stage (e.g. partial reading event) rather than the whole event. The culmination implicature comes from the fact the the event denoted by e' can be equal to the (whole) P-event denoted by e.

The semantics proposed, however, cannot explain the *konstatacija fakta* use of the Slovenian imperfective with achievements, since it predicts IPF_{SLO} to behave like IPF_{RU}, contrary to what is the case, cf. §2.1. Slovenian imperfective achievements have neither culmination entailment nor culmination implicature.[9]

[7] Rothstein [11] argues that achievements do not have stages in the first place, which is why coercion takes place. As [2] points out, this cannot account for the Russian data.

[8] Note that the type of data presented in this paper does not seem to vary with respect to whether the imperfective predicate is 'primary' or secondary (derived from a perfective verb), as shown by using examples of both. Since Altshuler's IPF_{RU} is a secondary imperfective, we will assume to be formalising the same kind of imperfective.

[9] If imperfective achievements had culmination implicatures in Slovenian, we could argue for underspecification: these predicates are not achievements per se, but are

Intuitively, it seems as if IPF_{SLO} was compatible only with durative events, such as activities or accomplishments, and not with near instantaneous events such as achievements.[10] We can formalise this intuition by appealing to the notion of atomic stage: IPF_{SLO} does not combine with atomic stages. Formally, I propose the following:

(12) $[\![IPF_{SLO}(P)(e')(w^*)]\!]^{\mathcal{M},g} = 1$ iff $[\![\exists e \exists w\ STAGE^S(e', e, w^*, w, P)]\!]^{\mathcal{M},g} = 1$ iff there is some g' that differs from g in at most the values for e and w such that:

a. the history of $g'(w)$ and $g'(w^*)$ is the same up to and including $\tau(g'(e'))$

b. $g'(w)$ is a reasonable option for $g'(e')$ in $g'(w^*)$

c. $[\![P(e)(w)]\!]^{\mathcal{M},g'} = 1$

d. $g'(e') \leq g'(e)$

e. $[\![\exists e'' \exists w''\ STAGE(e'', e, w'', w, P)]\!]^{\mathcal{M},g'} = 1$

Clause (e) essentially states that the event denoted by e must have at least one proper stage, from which it follows by definition that it is not atomic (so not an achievement). To enfold its semantics in an analogous way to (11): there must be an assignment function g'' that differs from g' in at most the values for e'' and w'' such that: (i) the history of $g''(w)$ and $g''(w'')$ is the same up to and including $\tau(g''(e''))$, (ii) $g''(w)$ is a reasonable option for $g''(e'')$ in $g''(w'')$, (iii) $[\![P(e)(w)]\!]^{\mathcal{M},g''} = 1$, and (iv) $g''(e'') < g''(e)$.[11]

The reader will have observed that clause (e) does not play a role in how the culmination implicature is derived with accomplishments, see further above, and that clause (e) is satisfied when the event denoted by e is an accomplishment since accomplishments have at least two stages.

On the other hand, clause (e) (more precisely, clause (iv)) leads to a clash with achievement predicates (since they do not have proper parts), and coercion must take place, as in English. In such cases, achievements are said to be restructured into 'abstract' accomplishments (they have the structure of an accomplishment, but they do not 'correspond' to any lexical item [11]). Roughly speaking, the preparatory stage becomes the 'process' part of the accomplishment and the

---

predicates specified only for telicity. See [2] for how the argument works for certain Russian predicates.

[10] The same observation is also found in [6].

[11] We could have formalised this idea as follows: $[\![IPF_{SLO}(P)(e')(w^*)]\!]^{\mathcal{M},g} = 1$ iff $[\![\exists e \exists w \exists e'' \exists w''(STAGE^*(e', e, w^*, w, P) \wedge STAGE(e'', e, w'', w, P))]\!]^{\mathcal{M},g} = 1$. This is perhaps a better illustration of how the Slovenian imperfective relates to the Russian imperfective and the English progressive. As noted in the outset in §1, it shares some of the relevant properties with both, though is neither completely. Nevertheless, as mentioned in footnote 3, the Russian and the Slovenian imperfective (unlike the English progressive) are compatible with states, which do not have stages. Therefore, the semantics of IPF_{SLO} will eventually have to get rid of the English STAGE, so the parallel works only for non-stative eventualities.

culmination point of the accomplishment is the achievement itself, see [11] for more details. In §2, example (7), we observed that the Slovenian imperfective and the English progressive focus on the interval before the culmination (e.g. the interval before reaching the top of a mountain), and that culmination is not implied. Given that the culmination implicature (with accomplishments) was argued to come about via the possibility of the event denoted by e' being equal to the event denoted by e, it is expected – due to the *strict* part-of requirement in (11d) – that there be no such implicature in English. The question is why such an implicature disappears with Slovenian 'restructured' accomplishments.[12]

The implicature disappears due to pragmatic reasons. Recall that achievements are one of the simplest event types – they consist of the culmination point only and do not have any subevents (proper stages). They are perfectly suited to combine with the perfective aspect to convey that this culmination point occurred. Recall also that the denotation of a Slovenian imperfective achievement is the result of a clash, which triggers coercion. So the perfective aspect is the only aspect that can naturally (i.e. without a clash) apply to an achievement eventuality. Since achievements are single-staged and consist of the culmination point only, why would a speaker use the imperfective aspect if not to avoid the inference that the achievement culminated? It seems uncooperative to trigger a clash and use this clash-induced imperfective achievement to also want to *imply* what is expressed with (and seems to be the sole purpose of) the only form that does not lead to a clash. Note that the same argument does not apply to accomplishments since imperfective accomplishments do not come about as the result of a clash.[13]

5 Conclusion

The previous section provided a formalisation of the Slovenian imperfective operator IPF_{SLO}, whose properties were discussed in §2. It was argued that IPF_{SLO} had the semantics of the Russian operator IPF_{RUS} with an additional constraint which accounted for its behaviour with achievement predicates.

In this concluding discussion, I wish to turn to the issue of imperfective performatives. In a recent article, Condoravdi and Lauer [4] propose that performative verbs 'denote communicative events [...] whose truth-conditional content is fully specified in terms of speaker commitments' [4, p. 13]. In order to derive the

[12] Suppose for a moment that coercion returned the preparatory process only (without the culmination point), and that this process consisted of several stages. There seems to be no empirical evidence to suggest that the English progressive (requiring a proper stage) focuses on a 'smaller' stage of the preparatory process than the Slovenian imperfective. This suggests that coercion really triggers the construction of an accomplishment, that the two operators can focus on the whole process before the culmination, and that IPF_{SLO} should be able to also denote the whole event, as with 'normal' accomplishments.

[13] Note also that the same argument does not apply to Russian achievements due to the semantics of the Russian imperfective, which does not lead to a clash.

performative effect, their theory requires the speaker to *have the commitment* to having a belief or an intention (he need not have the actual belief or intention) [4, p. 14]. They also predict the following:

(13) *Our account plus the assumption that performative verbs are accomplishments implies that the utterance of a performative progressive sentence does not commit the speaker to the existence of a commitment. This is so because progressive sentences describing accomplishments do* not *entail the culmination of the described event.* [4, p. 162]

In other words, a progressive sentence like (8b) does not commit the speaker to having the intention to come, which is necessary to derive performativity in [4]. The reason for this is that it describes an accomplishment in progress, and not one that has culminated. The Slovenian imperfective, on the other hand, can describe accomplishments whose culminations are *implied*, cf. §2.1. I leave it to further investigation whether this could be enough to derive performativity. Given how commonplace Slovenian imperfective performatives are, we could formulate an analysis on which the culmination is 'somehow' *required*, and this can be provided equally well by the English perfect as by the Slovenian imperfective, given the semantics put forward in this paper.[14]

[14] Russian also has imperfective performatives [5]. Note that there are also some, rather exceptional, cases of English progressive performatives [12, p. 537]. One could argue (Condoravdi, p.c.) that progressive performatives are not truly performative but are merely 'redescriptions' or restatings of what had already happened. (For instance, *I'm telling you he's an idiot* could be a mere restating of what I had already told you and not a performative utterance per se.)

References

1. Altshuler, D.: *Temporal Interpretation in Narrative Discourse and Event Internal Reference.* PhD thesis, New Brunswick (2010)
2. Altshuler, D.: There Is No Neutral Aspect. In Snider, T. (ed.) *Proceedings of SALT 23*, pp. 40–62 (2013)
3. Comrie, B.: *Aspect.* Cambridge University Press, Cambridge (1976)
4. Condoravdi, C. and Lauer, S.: Performative Verbs and Performative Acts. In Reich, I. et al. (eds.) *Proceedings of Sinn und Bedeutung 15*, pp. 1—15 (2011)
5. Dickey, S. M.: *Parameters of Slavic Aspect: A Cognitive Approach.* CSLI Publications, Stanford (2000)
6. Dickey, S. M.: Verbal Aspect in Slovene. *Sprachtypologie und Universalienforschung* 56(3), 182–207 (2003)
7. Filip, H. and Rothstein, S.: Telicity as a Semantic Parameter. In Lavine, J. et al. (eds.) *The Princeton University Meeting*, pp. 139–156, Michigan Slavic Publications, Ann Arbor, MI (2005)
8. Glovinskaja, M.: Semantika, pragmatika, i stilistika vido-vremennyx form. In Šmelev, D. N. (ed.) *Grammatičeskie issledovanija.* Nauka, Moscow (1989) [cited in [5]]
9. Landman, F.: The Progressive. *Natural Language Semantics* 1, 1–32 (1992)
10. Orešnik, J.: *Slovenski glagolski vid in univerzalna slovnica.* SAZU, Ljubljana (1994)
11. Rothstein, S.: *Structuring Events: A Study in the Semantics of Lexical Aspect.* Blackwell, Oxford (2004)
12. Searle, J. R.: How Performatives Work. *Linguistics and Philosophy* 12(5), 535–558 (1989)
13. Smith, C. S.: *The Parameter of Aspect.* 2nd edition. Kluwer, Dordrecht (1997)
14. Toporišič, J.: *Slovenska slovnica.* Založba Obzorja, Maribor (2000)
15. Žagar, I. Ž.: Performativity as Tense and Aspect. *International Review of Pragmatics* 3(2), 168–193 (2011)

Unless, Exceptionality, and Conditional Strengthening

Prerna Nadathur

Department of Linguistics, Stanford University, California

Abstract This paper discusses several challenges to the exceptive account of the connective *unless* ([4], [5], [15]). I argue that exceptionality fails to capture the truth conditions for *unless* under non-universal quantifiers (e.g. *most*) and also does not account for the infelicity of *unless*-conditionals in certain circumstances. I propose instead that *unless* shares asserted content with *if not,* and its apparent biconditionality (under positive quantifiers) is due to a generalized conversational implicature akin to conditional perfection ([8]). Further, I propose that statements of the form q COND p are subject to a felicity inference that the speaker is unwilling/unable to assert q unconditionally: this is an implicature for regular *if*-conditionals, but apparently presuppositional for *unless*. This difference accounts for the divergent pragmatics of *unless* and *if not*.

Keywords: unless, conditionals, exceptive constructions, conditional perfection, implicature, presupposition

1 Introduction

Higginbotham [11] puts forward the subordinating conjunctions *if* and *unless* as putative counterexamples to a compositional theory of semantics, claiming that they vary in their contribution when embedded under different quantifiers. Although various "fixes" have been proposed for this problem (e.g. [19]), these have proven unsatisfactory enough that the *unless* counterexample, at least, has entered the literature as a standard objection to compositionality (e.g. [12], [23]). A resolution of Higginbotham's puzzle therefore holds a certain significance for the debate over compositionality at large.

With respect to the logic of natural language, *unless* is also of interest as an apparently propositional operator which cannot be accounted for on purely truth-functional grounds. In this paper, I first discuss the challenges to a truth-functional account, and then argue that even the prominent "exceptive" treatment of *unless* (see [4], [5], [15]) fails in several respects. First, it is too strong in positive quantificational contexts. Second, it fails to capture the truth conditions for *unless* under non-universal quantifiers (such as *most*). Third, it does not account for situations in which the use of *unless* seems infelicitous. To handle these considerations, I propose that *unless* shares assertive content with *if not*, and argue that the difference between the two conditionals is located in their

association with a specific "conditional strengthening" inference, which for *unless* is presuppositional and for *if not* is a conversational implicature. This raises several interesting new empirical questions, which I outline in the final sections.

2 Truth conditions and Higginbotham's puzzle

Classically, *unless* is treated as equivalent to the negative material implication *if not* (q *unless* $p \leftrightarrow q$ *if not* p), which reduces to regular disjunction ([21], [22]):

(1) q **unless** $p := \neg p \rightarrow q \qquad (\leftrightarrow p \vee q)$

This seems at least acceptable in "bare" or positively-quantified contexts (although more recent work, such as [4] and [15], suggests that unidirectionality is weaker than intuition requires here; this is discussed in detail below):

(2) a. John will leave unless Bill calls $= \text{call}(B, J) \vee \text{leave}(J)$
b. Everyone will leave unless Bill calls them $= \forall x\ (\text{call}(B, x) \vee \text{leave}(x))$

However, negative material implication (or disjunction) supplies the wrong truth conditions under negative quantifiers. For instance, in (3), it would require that each person neither leaves nor is called by Bill.

(3) No one will leave unless Bill calls them $= \neg\exists x\ (\text{call}(B, x) \vee \text{leave}(x))$
$= \forall x\ (\neg\text{call}(B, x) \wedge \neg\text{leave}(x))$

Higginbotham argues that an intuitively acceptable interpretation for (3) is given by replacing *unless* with *and not*: thus, $\neg\exists x\ (\text{call}(B, x) \wedge \neg\text{leave}(x))$. On this basis, he concludes that *unless* varies in meaning according to the context in which it appears, and attributes to it a noncompositional semantics.

As noted, there is reason to question equating *unless* with material *if not* even in the positive cases. Other truth-functional proposals in the literature include *not p only if q* (see [3]) and *q only if not p* (see [1]). Together, these give the biconditional $[\neg p \leftrightarrow q]$, which seems to better capture the strength of an *unless*-statement:

(2′) a. Everyone will leave unless Bill calls them $= \forall x\ \neg\text{call}(B, x) \leftrightarrow \text{leave}(x)$

Unfortunately, the biconditional too provides the wrong interpretation in negative contexts. The intuitive interpretation of (3) holds that no one who is not called by Bill leaves, but does not necessarily stipulate that all those called by Bill do leave. That is, Bill calling should only be a necessary condition, but not a sufficient one – it should be possible that some people who are called by Bill do not leave. Higginbotham's compositionality puzzle apparently persists, but in a modified form: we seem to want a biconditional *unless* in positive contexts, but a unidirectional one in negative contexts.

3 Exceptionality

3.1 Uniqueness and biconditionality

The most current account of *unless* treats it as an exceptive operator. The central idea is that *unless*-statements assert a generalization, and in addition assert the existence of an exception to that generalization (cf. [7], [2]). More specifically, *unless* can only occur in the scope of a quantifier; it subtracts from the domain over which the quantified statement is evaluated, and asserts that the complement of *unless* represents an exception to the quantified statement. The first formalization of this is due to von Fintel [4], [5], who argues that the *unless*-complement marks the *unique* smallest exception to the quantified proposition. This gets around Higginbotham's original compositionality puzzle by replacing material implication with the restrictive Lewis-Kratzer conditional ([17], [13]), but unfortunately produces an essentially biconditional interpretation in negative contexts. Leslie [15] refines von Fintel's proposal to address the biconditionality/unidirectionality puzzle, proposing (4) as the semantics of a quantified *unless*-statement. Square brackets enclose the quantifier's restriction.

(4) $Q[C]M$ **unless** $R := Q[C \wedge \neg R]M \wedge Q[C \wedge M]\neg R$

In (4), Q is a quantifier (or quantificational adverb), C is the domain of the quantifier (for instance, a set of relevant situations or possible worlds), M is its nuclear scope, and R is the *unless*-complement or excepted set. The first conjunct is the *if not* direction, while the second is *not if* ("uniqueness"; [4], [5]). This gives biconditionality for (2)a, but interprets (3) as follows:

(3′) No one will leave unless Bill calls them
$= \neg\exists x[\text{per}(x) \wedge \neg\text{call}(B, x)]\ \text{leave}(x)\ \wedge\ \neg\exists x[\text{per}(x) \wedge \text{leave}(x)]\ \neg\text{call}(B, x)$
$= \neg\exists x[\text{per}(x) \wedge \neg\text{call}(B, x)]\ \text{leave}(x)$

Formula (4) exploits the symmetric nature of the negative universal quantifier to reduce both conjuncts to equivalent statements. Biconditionality therefore evaporates in (3′), and we get only that no one leaves who is not called.

3.2 Other determiners

Even (4) runs into trouble when considered against non-universal quantifiers such as *most*, *some*, and *half*. The naturally-occurring examples in (5) show that *unless* is not limited to universal contexts, as claimed by von Fintel in [4].

(5) a. "Most in the U.S. support a higher minimum wage, unless it costs jobs."
b. "Some diners won't get water unless they ask."
c. "Smoking kills half of smokers unless they quit."

Each of these is problematic for (4). Consider the following illustration:

(6) Most students will succeed unless they goof off
= Most x [st(x) $\wedge$ $\neg$goof(x)] succ(x) $\wedge$ Most x [st(x) $\wedge$ succ(x)] $\neg$goof(x)

Suppose our universe is a class of 12 students, only 4 of whom goof off. 6 of the non-goofing students succeed, and 3 of the others do too. Then we have 6 successful students from the non-goofing 8, so the first conjunct is satisfied. We also have 6 non-goofing students of 9 successful ones, so the second conjunct is satisfied. The problem is that most of the students who do goof off also succeed – goofing off makes no difference to the success rate (most succeed no matter what). (6) seems an inappropriate statement here, but (4) fails to capture this.

Finally, it is also not obvious that we want unidirectionality only in negative contexts. *Unless* does not mandate a biconditional interpretation for (7)a; this patterns instead with (3), in that Mantou being out is necessary to prevent her lateness, but may not be sufficient. (7)b corroborates this.

(7) a. "Mantou is always late unless she's already out before we meet."
b. "Mantou is always late unless she's already out before we meet, but she's often just less late then."

If *unless* were truly biconditional here, (7)b should be as contradictory as "Roses are always red and violets are always blue, but violets are not always blue."

4 The pragmatics and semantics of *unless*

4.1 Uniqueness as a pragmatic inference

The preceding discussion suggests that, while the exceptive treatment correctly identifies both *if not* and *not if* directions as relevant to the interpretation of *unless*, it goes wrong in attributing to them the same status. In particular, (7) argues against treating the *not if* direction as entailed or asserted content. A few additional observations support the claim that uniqueness is pragmatic.

First, the natural example (8) shows that uniqueness can be reinforced without redundancy. Entailments do not have this property.

(8) "Always be yourself, unless you are Fernando Torres. Then always be someone else."
Compare: Always be yourself, unless you are Fernando Torres. ?Otherwise always be yourself.

Uniqueness here gives something like "Always be not yourself, if you are Fernando Torres." The second sentence therefore represents a grammatically standard reinforcement of uniqueness. Crucially, this does not appear redundant in the way that the comparison's reinforcement of the *if not* entailment does.

Second, while it is contradictory to question an entailment, the natural example in (9) shows that uniqueness *can* be questioned (since it would give here that the answer cannot be no if you do ask):

(9) "The answer is no unless you ask. If you do ask the answer might be no."
Compare: The answer is no unless you ask. #If you don't ask the answer might be yes.

Finally, uniqueness in (7)a would hold that Mantou is never late when she is already out, so (7)b shows that uniqueness is defeasible. In particular, the inference of biconditionality is cancelled by the assertion that Mantou is late in at least some of the situations where she is out beforehand. Defeasibility creates a heavy presumption in favour of a pragmatic account for uniqueness.

Descriptively, uniqueness can be classified as a generalized conversational implicature (GCI) *à la* Levinson [16]. (10)a suggests that it is not presuppositional, since it can be suspended prior to an *unless*-statement without causing infelicity, and (10)b shows that it is not redundant when backgrounded, which goes against a conventional implicature treatment (see [20]).[1]

(10) a. The student might not fail if he studies, but he'll fail unless he studies.
Compare: ?There might not be a student, but the student will fail unless he studies.
b. John won't fail if he studies. He will fail unless he studies.
Compare: John is a student. John, ?the student, will fail unless he studies.

Finally, the regularity of the inference to uniqueness (as attested by attempts to capture it in a semantic treatment) stands against classifying it as a particularized conversational implicature. Instead, it "captures our intuitions about preferred or normal interpretations" ([16], p.11).

Uniqueness bears a strong resemblance to another GCI, conditional perfection (the biconditional interpretation often attributed to *if*-conditionals; see [8]). Both inferences are defeasible, reinforceable, and nonconventional (in the sense that they are noncoded and do not attach in all circumstances, e.g. negative contexts for *unless*). Both may be seen as instantiations of Levinson's I-heuristic (which he ties to Grice's Quantity-2 maxim for the speaker).[2] The default nature of uniqueness is what makes it appear so ubiquitous when it does attach.

4.2 The semantics of *unless*

Having consigned the *not if* direction to the pragmatics, we have only *if not* left for the semantics. The exceptive account provides the tools for avoiding the

[1] The examples in (10) may prefer prosodic focus on *unless* due to the parallelism between the clauses, but I do not believe this is essential to their acceptability.

[2] The I-heuristic directs the listener to "[maximize] informational load by narrowing the interpretation to a specific subcase of what has been said" ([16], p.118).

compositionality problem originally encountered by Higginbotham in trying to equate *if not* with *unless*; here, I spell out the essential pieces of this treatment.

In particular, I claim that the first conjunct of formula (4) gives us precisely the asserted content of *unless*. This conjunct is common to both the von Fintel and Leslie accounts, and crucially incorporates the Lewis-Krazter restrictive conditional ([17], [13]). A "bare" *unless*-statement, *q unless p*, is therefore presumed to contain a covert universal modal *must*, as in (11).

(11) must q **unless** $p := \forall w[\neg p(w)]\ q(w))$

Furthermore, Leslie argues that it is insufficient to regard conditional operators as only capable of restricting modal quantifiers, and extends the restrictor analysis to quantificational determiners as well.[3] Absent this modification, we continue to get the wrong interpretation for examples like (12), which contain overt quantificational determiners. (12)b gives the faulty interpretation; compare this to (12)c, which incorporates Leslie's version of the restrictive conditional.[4]

(12) a. No one will leave unless Bill calls them.
b. **Lewis-Kratzer:** $\neg\exists x\ (\forall w[\neg\text{call}(B, x, w)]\ \text{leave}(x, w))$
c. **Modalized Restrictor (Leslie):** $\forall w\ (\neg\exists x[\neg\text{call}(B, x, w)]\ \text{leave}(x, w))$

(12)b is logically equivalent to the following:

(12′) b. $\forall x\ (\exists w[\neg\text{call}(B, x, w)]\ \neg\text{leave}(x, w))$

Here, all individuals are such that there is some situation where Bill does not call and they do not leave. This is incorrect: it should be *all* not-calling situations in which they do not leave. (12)c avoids this problem by having *unless* restrict the nominal quantifier.

In summary, *unless*, like *if* and *if not*, is a restrictive operator on quantifiers. It imposes the same restriction as *if not*: the negation of its complement proposition. This restriction is imposed on the covert universal modal in a bare conditional, but when more than one quantifier is present (as in (12), where the quantificational determiner is overt and the modal is covert), *unless* restricts the quantifier with narrower scope (subject to the usual ambiguities) and scopes under the other (see [9] for a more detailed examination of these interactions).[5] (13) shows how this works for a conditional with two overt quantifiers.

[3] This is an extension with respect to [13], but is in the spirit of [17].

[4] Leslie [15] provides critical arguments for the wide-scope universal modal. I do not discuss these here, but my account fully adopts Leslie's "modalization."

[5] There are also cases, such as "Most people go swimming outside unless it's raining" where the overt quantifier does not seem to scope over the p clause, and here the conditional must restrict the covert modal. These can also be handled by the outlined account, although I do not discuss this here. See also [9].

(13) a. No one usually leaves unless Bill calls them.
b. $\neg\exists x$ (Most $w[\neg\text{call}(B, x, w)]$ leave(x, w))
c. Most w ($\neg\exists x[\neg\text{call}(B, x, w)]$ leave(x, w))

(13)b provides the reading on which the quantifiers scope in the order they appear: no one is such that most non-calling situations are ones in which they leave. (13)c is the alternative: most situations are such that no one leaves who has not been called by Bill.

Leslie's modalized restrictor account can thus provide a semantics for *unless* as well as *if not*. Crucially, it is invariant to quantificational context, and so avoids the compositionality problem.

5 Chasing the difference between *unless* and *if not*

Section 4 leaves us with a new problem: if *unless* and *if not* share asserted content, why are they pragmatically different? (14)a and b seem to differ precisely in the extent to which they invite uniqueness: very strongly and relatively weakly (absent a suggestive context), respectively.

(14) a. John will leave unless Bill calls him.
b. John will leave if Bill does not call him.

Why does *unless* default to biconditionality here when *if not* does not?[6]

5.1 Conditional strengthening

The first step in explicating the difference is to observe that conditionals are accompanied by a certain presumption about the circumstances in which they can be felicitously uttered. Following von Fintel [6], who remarks on a similar inference, I refer to this as "conditional strengthening."

(15) **Conditional Strengthening:** Given a conditional operator COND and two propositions p and q, the statement q COND p is "best" asserted when the speaker is unable/unwilling to assert the unqualified proposition q.

This is relatively uncontroversial: it is peculiar to use a conditional if you could have simply asserted its consequent. Examples like (12) show that the strengthening proposition does not simply make reference to the proposition q, but "reaches up" to (at least) the restricted quantifier.

(12) a. No one will leave unless Bill calls them.
$\rightsquigarrow$ The speaker is unwilling/unable to assert "No one will leave."

[6] This is particularly relevant since both Grice [10] and Levinson [16] expect conversational implicatures to display nondetachability.

[6] treats conditional strengthening as an implicature,[7] but it has some special properties. For one, it does not seem to be defeasible in the usual sense:

(16) ?John will leave if Bill does not call. Actually, he will leave no matter what.

(16) is peculiar because it is difficult to construct any excuse for the use of a conditional statement when the speaker is willing and able to commit to the unconditional proposition q. This is essentially the logic behind Lauer's [14] "Need a Reason" (NaR) implicatures: a proposed class of conversational implicatures that are crucially neither optional nor defeasible. The motivating example in [14] involves the "ignorance" implicature associated with disjunctive statements:

(17) John is in Paris or he is in London.
$\rightsquigarrow$ The speaker is unable/unwilling to say which.

Paraphrasing from [14], a general communicative preference for less complex utterances can only be overridden if there is a reason for the speaker to do so; in the case of (17), and with conditionals, the reason cannot be informativity, since the shorter alternative is actually more informative. The reason must therefore be something else; but the *need* for a reason is, crucially, what cannot be canceled. This is motivated in greater detail by [14].

5.2 Conditional strengthening for *if not* and *unless*

Although conditional strengthening is associated with both *if* and *unless*, it is only with *if*-conditionals that it represents an NaR implicature. It turns out to be stronger with *unless*. The critical observation is that, while NaR implicatures are non-defeasible, they remain implicatures and cannot affect truth conditions.

Suppose you are presented with a box of red marbles. (18b) is uncontroversially a better description of this situation than (18)a. Nevertheless, (18)a is *true*, even in the absence of a contextually recoverable reason for choosing the less simple alternative.

(18) a. Every marble is red or blue.
b. Every marble is red.

Conditional strengthening with *if not* patterns the same way. Suppose half of the marbles in the box are red, and the other half are blue, and suppose further that every marble, regardless of colour, has a black dot. Here, (19)a is again uncontroversially true, despite the fact that (19)b is a better description of the situation, and even though there is no recoverable reason for asserting (19)a instead of (19)b. On the other hand, (19)c is false or otherwise noticeably infelicitous, and this seems to be because (19)b is observably true. Nadathur and

[7] In particular, as a scalar implicature from the scale $\langle$whatever the case, $q >$ if p, $q\rangle$.

Lassiter (to appear; [18]) provide experimental verification of these predictions.

(19) a. Every marble has a dot if it is not blue.
b. Every marble has a dot.
c. Every marble has a dot unless it is blue.

The examples in (19) show that, while conditional strengthening on *if not* is "suspendable" (in a context-free, forced-choice scenario), *unless*-conditionals cannot be used without it. This is corroborated by (7)b and (9): in canceling the inference to biconditionality, both examples avoided denying conditional strengthening in order to provide an acceptable statement. Compare the following:

(7′) b. #Mantou is always late unless she's already out before we meet. She's always late no matter what.

(9′) #The answer is no unless you ask. If you do ask the answer is no.

The false judgement for (19)c seems to me on par with a description of "The King of France is bald" as false (which is a judgement often given by those encountering the example for the first time). In particular, (19)c is "false" in the sense that the assertion is infelicitous, because conditional strengthening (that the speaker cannot say (19)b, in this case) is defeated. This suggests that strengthening is a presupposition associated with the use of *unless*. This classification no doubt warrants closer examination, but the crucial observation is that conditional strengthening, which I argued to be an (NaR) implicature for *if* and *if not*, is an altogether stronger proposition with *unless*. This accounts for the pragmatic divide between *if not* and *unless*, as seen in (14) and (19).

Finally, consider one additional scenario. We again have a box of marbles, half of which are blue, and the other half red. In this case, however, none of the marbles has a dot. Again, the *if not* statement (20)a is true, but the *unless* statement (20)c is judged false (or infelicitous), due to the availability of the unconditional alternative (20)b. See [18] for the corroborating experimental data.

(20) a. No marble has a dot if it is not blue.
b. No marble has a dot.
c. No marble has a dot unless it is blue.

The account presented here correctly predicts these results, and in particular the false judgement for (20)c. According to the exceptive account in formula (4), however, *unless* in the negative context is identical to *if not*, and (20)c is therefore predicted to be true.[8] This is a critical difference between the two accounts.

[8] Von Fintel's original formulation of the exceptive account has biconditionality here, which correctly predicts falsity. However, this formulation gets the case where some but not all of the blue marbles have dots wrong: biconditionality forces this to be false, while the account developed here (correctly) predicts (19)c to be true.

In addition to this, conditional strengthening captures precisely what goes wrong with the exceptive account for examples like (6).

(6) Most students will succeed unless they goof off.

Since, in the scenario presented in 3.2, most students succeed no matter what, conditional strengthening correctly predicts that (6) is an infelicitous statement, where formula (4) does not.

The exceptive treatment is too stringent in the positive cases and too lenient in the negative ones. Recognizing first that uniqueness is a GCI and, secondly, that conditional strengthening is an implicature with *if not* but a presupposition with *unless* allows us to split the difference and capture the empirical data.

5.3 Conditional strengthening and biconditionality

Often, the most immediate inference from a conditional is that the speaker's reason for not asserting the simpler statement q is that she knows q not to be universally true (or at least lacks evidence for this claim). One way of restating this is to say that a conditional utterance q *unless* p often (pre)supposes that there are (epistemically) relevant situations such that $\neg q$.

(21) John will leave unless Bill calls.
Assertion $\forall w[\neg\text{call}(B, J, w)]\ \text{leave}(J, w)$
Presupposition $\exists w\ \neg\text{leave}(J, w)$

Consider again what a conditional like (21) does. We are offered two ways of partitioning the set of relevant situations: first on the basis of whether Bill does or does not call, and secondly on the basis of whether John does or does not leave. Each forms a partition of the whole space. Minimally, (21) requires that the set of situations in which Bill does not call form a subset of the situations in which John leaves (this is given by the asserted content of *unless/if not*). This does not tell us anything about the partition associated with whether or not John leaves: it could be trivial (John leaves in all situations) or nontrivial (there exist situations in which John does not leave). As a presupposition, conditional strengthening fixes that we are in the nontrivial case.

It is a logical consequence of the semantics of *unless* that the set of situations in which John does not leave forms a subset of the set of situations in which Bill calls. As stated, conditional strengthening tells us that this subset is nonempty – there are situations in which John does not leave. It therefore provides something like a "foothold" for the inference to full conditional perfection, which is the conclusion that the set of situations where John does not leave and the set of situations in which Bill calls are *coextensive*. This is the "best-case" (or most appropriate) scenario for the use of an *unless*-conditional and is taken as a default interpretation in many contexts.

Since conditional strengthening is a precondition for the use of an *unless*-conditional, we are always in the "nontrivial" case with *unless*. The suggestion I

am making here is that this lends itself to perfection/uniqueness, which is thus very strongly associated with *unless*. Although *if not* conditionals are typically also inferred to be describing the nontrivial case, the decreased stringency of the strengthening inference here carries over to a decreased association with perfection. The details of this suggestion, of course, warrant further investigation.

6 Summary and Outlook

Starting from Higginbotham's observations, I have presented and discussed the challenges faced in developing a compositional account of *unless*. I have argued that the prominent exceptive solution to these challenges is insufficiently sensitive to the distinction between the asserted and non-asserted content associated with *unless*. In particular, I have shown that the two conditional "directions" in fact belong to different classes of meaning: the *if not* conditional, formulated as per Leslie's modalized restrictor treatment, fully captures the asserted content, and the *if not* conditional is a GCI akin to conditional perfection.

In addition, I have argued that the difference between *if not* and *unless* conditionals is due to the difference in their association with the (felicity) inference of conditional strengthening. This is an NaR implicature (see [14]) for *if not*, but is stronger for *unless*. Insofar as conditional strengthening provides a foothold for the uniqueness (*not if*) implicature, this may account for the difference between *if not* and *unless* with respect to the appearance of biconditionality.

A number of questions remain open, and offer interesting avenues for further investigation. First, as noted in section 5.2, [18] provides experimental evidence supporting the claims surrounding examples (19)-(20), and my further investigation involving the non-universal quantifiers *most, some* and *few* so far suggests that the account presented here can address the shortcomings of the exceptive treatment with respect to these quantifiers as well. Second, experimental manipulation of the context provided as a background for *if not*- and *unless*-statements is likely to shed light on the precise details of the inference to conditional strengthening. Third, it remains to be seen whether the basis provided here for differentiating between *if not* and *unless* is sufficient to explain why uniqueness is so ubiquitous in positive contexts and so noticeably absent in negative ones. I suspect that the difference can be related to how strongly these (quantificational) contexts support the foothold that strengthening provides for perfection, but this is yet to be thoroughly investigated. Finally, insofar as [4], [5], and [15] (among others) treat *unless* as belonging to a class of exceptive operators, it seems worth exploring whether other members of this class can also be handled by proposals like the one presented here, and in particular whether their behaviour can also be fruitfully explicated as a consequence of association with specific non-defeasible inferences.

References

1. Clark, H., and Clark, E. *Psychology and Language: An Introduction to Psycholinguistics.* Thomson Learning, London (1977)
2. Dancygier, B. If, unless, and their Polish equivalents. *Papers and Studies in Contrastive Linguistics* 20, 64–72 (1975)
3. Fillenbaum, S. The use of conditionals in inducements and deterrents. In E. Traugott, A. ter Meulen, J. Reilly, C. Ferguson (eds.), *On Conditionals*, 179–195. Cambridge University Press, Cambridge (1986)
4. von Fintel, K. Exceptive conditionals: the meaning of "unless". *Proceedings of the North East Linguistics Society* 22, 135–151 (1992)
5. von Fintel, K. Restrictions on quantifier domains. PhD thesis, University of Massachusetts–Amherst (1994)
6. von Fintel, K. Conditional strengthening: a case study in implicature. Ms, MIT (2001)
7. Geis, M. "If" and "unless". In B. Kachru, R. Lees, Y. Malkiel, A. Pietrangeli, S. Saporta (eds.), *Issues in Linguistics: Papers in Honor of Henry and Renee Kahane,* 231–253. University of Illinois Press, Urbana IL (1973)
8. Geis, M. and Zwicky, A. On invited inferences. *Linguistic Inquiry* 2, 561–566 (1971)
9. Geurts, B. On an ambiguity in quantified conditionals. Ms, University of Nijmegen (2004)
10. Grice, P. *Studies in the Way of Words.* Harvard University Press, Cambridge MA (1975)
11. Higginbotham, J. Linguistic theory and Davidson's program in semantics. In E. Lepore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson,* 29–48. Blackwell, Oxford (1986)
12. Janssen, T. Compositionality. In J. van Benthem, A. ter Meulen (eds.), *Handbook of Logic and Language,* 417–475. Elsevier, Amsterdam (1997)
13. Kratzer, A. Conditionals. *Proceedings of the Chicago Linguistic Society* 22, 1–15 (1986)
14. Lauer, S. Towards a dynamic pragmatics. PhD thesis, Stanford University (2013)
15. Leslie, S.-J. "If," "unless," and quantification. In R. Stainton, C. Viger (eds.), *Compositionality, Context, and Semantic Value: Essays in honor of Ernie Lepore,* 3–30. Springer, Amsterdam (2008)
16. Levinson, S. *Presumptive Meanings: the Theory of Generalized Conversational Implicature.* MIT Press, Cambridge MA (2000)
17. Lewis, D. Adverbs of quantification. In P. Portner, B. Partee (eds.), *Formal Semantics: the Essential Readings,* 178–188. Blackwell, Oxford (1975)
18. Nadathur, P. and Lassiter, D. Unless: an experimental approach. *Proceedings of Sinn und Bedeutung* 19 (2014; to appear)
19. Pelletier, F. On an argument against semantic compositionality. In D. Prawiz, D. Westerstahl (eds.), *Logic and Philosophy of Science in Uppsala,* 599-610. Kluwer, Dordrecht (1994)
20. Potts, C. *The Logic of Conventional Implicatures.* Oxford University Press, Oxford (2005)
21. Quine, W. *Methods of Logic.* Holt, Rinehart & Winston, New York (1959)
22. Reichenbach, H. *Elements of Symbolic Logic.* The Free Press, New York (1947)
23. Szabó, Z. Compositionality. In *The Stanford Encyclopedia of Philosophy* (2008)

The Different Readings of *Wieder* and *Again* - An Experimental Investigation

Anthea Schöller*

University of Tübingen, Germany
anthea.schoeller@uni-tuebingen.de

Abstract This paper experimentally examines which readings (repetitive/restitutive/counterdirectional) of the adverb *again/wieder* are still in common usage in English and German. The diachronic change can be explained with Rapp&Stechows (1999) Visibility Parameter for adverbs. Experiment 1 confirms the hypothesis that English *again* is moving from setting (iii) to (ii). This means that it is losing its ability to see into the components of a lexical accomplishment predicate and is becoming less and less common when having a restitutive meaning. German *wieder* in combination with a LA is still accepted in a restitutive context; it still has setting (iii). Experiment 2 shows that neither *again* nor *wieder* are used with a counterdirectional meaning any more. Two different semantic analyses are necessary to explain this diachronic development, the lexical analysis and the structural analysis.

Keywords: again, wieder, decomposition adverb, diachronic, experiment, lexical accomplishment

1 Overview

This paper presents empirical evidence which shows that English *again* is losing its restitutive reading and is shifting towards the repetitive, whereas both are still accepted in German. Furthermore, the results suggest that the counterdirectional reading has nearly disappeared in both languages. Two different semantic analyses are necessary to explain this.

2 Background

Sentences with the adverb *again* can be ambiguous (cf. McCawley 1968). In (1a) the whole event is repeated. In (1b) only the result state of the event is restored.

(1) Felix opened the window again.
   a) Felix opened the window and he had done that before. **repetitive**
   b) Felix opened the window and the window had been open before.
   . **restitutive**

* I thank Sonja Tiemann and Sigrid Beck for insightful comments.

Examples of counterdirectional (presupposing an action in a reversed direction) *again* from the literature are given in (2).

(2) "talk to them again" = reply to them; "write again to him" = write back to him

There are two competing approaches to analysing these sentences:
1) the **lexical analysis** (Fabricius-Hansen 2001)
2) the **structural analysis** (Stechow 1995).

In the lexical analysis the adverb has two different lexical entries, a repetitive (3)a and a counterdirectional (3)b. The analysis relies on conceptual prerequisites, such as the availability of a counterdirectional predicate, a result state and a prestate of an event.

(3) (a) [[again rep]] = λP. λe. $\exists$ e'[e'<e & P(e')]. P(e)
(b) [[again rest]] = λP. λe. $\exists$ e'[e'<e & P_c(e') & res_{Pc}(e') = pre_P(e)]. P(e)

In the structural analysis *again* always indicates repetition and only has the lexical entry (3)a. One must distinguish complex predicates, which have an overt result state (as in: Sonja painted the door blue.), from **lexical accomplishment predicates** (LA) as in ((1)), whose result state is not overt and can be accessed only after a process of decomposition. The adjunction sites of *again* bring about the different interpretations.

(4) (a) $\left[_{VP}\ [_{VP}$ Felix [$\emptyset_V$ [$_{SC}$ $open_{Adj}$ [the window]]]] again$\right]$ **repetitive**

(b) [$_{VP}$ Felix [$\emptyset_V$ $\left[_{SC}\ [_{SC}$ $open_{Adj}$ [the window]] again$\right]$]] **restitutive**

Rapp&Stechow (1999) propose a **Visibility Parameter for Adverbs** that Beck refined in 2005. An adverb like *again* is able to attach to a whole VP, to the result state of a resultative and can even see into the internal compositions of an accomplishment predicate. That is why these adverbs are called decomposition adverbs.

(5) The visibility parameter for adverbs (Beck 2005)
An adverb can modify
(i) only independent syntactic phrases
(ii) any phrase with a phonetically overt head
(iii) any phrase
The default setting is (i).

Being a decomposition adverb, *again* can modify any phrase. That means that theoretically the participants should be able to understand a restitutive reading in sentences with complex predicates. If they do not, a plausible explanation is that *again* with setting (iii) of the visibility parameter vanishes from their lexicon and moves to setting (ii) as proposed in Beck, Berezovskaya & Pflugfelder.

3 The puzzle

Which readings of *again/wieder* are still in common usage and how are they accounted for? The literature does not provide a consistent answer. In 2005, Beck described *again* and *wieder* as adverbs that have setting (iii). In a corpus study in 2009, Beck et al. found that the use of restitutive *again* is disappearing and that the repetitive meaning has become the most prominent. Not only the number of restitutive *again*s has diminished but also the range of predicates used with it. Especially LA are hardly combined with a restitutive *again* any more. This supports the structural analysis and suggests that *again* is moving from setting (iii) to (ii). In 2012 Gergel & Beck examined the transition from EModE to LModE. Their observation was that *again* was used in contexts where it did not have a repetitional meaning but was counterdirectional. Gergel & Beck propose that EModE has counterdirectional *again*; a counterdirectional analysis needs to be considered.

Hypotheses Based on the literature the hypothesis is that *again* has lost setting (iii). English speakers are expected to refuse a restitutive use of *again* with LA predicates, so (1a) should be the prominent reading. For German, the transition in the setting of *wieder* is expected to not have evolved that far; the test sentences in German should be better accepted. The rep./rest. ambiguity is accounted for by the structural analysis. To show that counterdirectional *again* does not exist any more in English and hardly in German is the objective of experiment 2. The counterdirectional meaning (as in 2) has disappeared from the lexicon. This requires the lexical analysis. All in all, both analyses are necessary to grasp the diachronic development.

4 Experimental setup

The study has a 2x2 design with fixed factors context (rep./rest.) and language (English/German). An online questionnaire was filled out in English or German. The sentences' acceptability was rated on a 4-point scale from 1 (acceptable) to 4 (not acceptable). There were 99 participants, 52 native English and 47 native German speakers.

4.1 Experiment 1

In the twelve target sentences (as in (6)) the adverb either has a repetitive or a restitutive reading, depending on the context introducing it.

(6) **Repetitive context in English** Jack is at primary school and is fascinated by pirates. In the playground he buried a necklace and marked the spot on a treasure map. Last month he dug out his treasure to make sure that it was still there. **He carefully examined the necklace and then buried it again.**

Restitutive context in English Jack is at primary school and is fascinated by pirates. In the playground somebody lost a golden necklace which was covered by a thick layer of leaves last autumn. Looking for treasures he accidentally found it and was thrilled about his first discovery. **He carefully examined the necklace and then buried it again.**

4.2 Experiment 2

Six sentences as in (7) with counterdirectional *again* or *wieder* were provided.

(7) In a newspaper advert Margarete was looking for a pen pal. A woman answered and sent her a nice letter. **Enthusiastically, Margarete wrote to her again.**

4.3 Results for restitutive *again*

The restitutive sentences are rated at a grand mean of 2.09 in English. This means that they are rather acceptable, but not as good as in German. With a 1.59 rating, *again*in restitutive sentences is acceptable. There is a significant interaction for the factor language. For **English**, the grand mean rating of the repetitive sentences is 1.88 and it is 2.09 for the restitutive sentences. The statistical analysis of the English sentences with an ANOVA reveals that the difference is significant ($p < 0.05$). For **German**, the means of repetitive and restitutive sentences do not differ significantly ($p > 0.1$). Both readings are rated roughly equally.

4.4 Results for counterdirectional *again*

The counterdirectional sentences are generally not accepted in both languages with a grand mean rating of 2.37 in English and 2.67 in German. The German speakers disapprove of every item with ratings of 2.28 or above and about one third did not accept any item.

5 Discussion

The results of the study in large support the initial hypotheses. The ratings for **restitutive** *again* and *wieder* suggest that in German the adverb is still accepted in combination with lexical accomplishment predicates and can attach to the result state after the decomposition process. English *again* is losing this feature. In other words, German *wieder* still has setting (iii) of the Visibility Parameter, but English *again* is moving from (iii) to (ii), its meaning is mainly repetitive; the structural analysis explains this nicely.

The **counterdirectional** reading of the adverb is not in common usage any more in either language; in German it is rejected even more firmly than in English. An interesting observation is that not even the complex predicate

anrufen is accepted in German. This supports Gergel & Beck's theory that counterdirectionality cannot be explained with the structural approach. Even with a separable verb counterdirectional *wieder* is not accepted. The structural approach explains sentence's ambiguity and the restitutive reading with the fact that the adverb decomposes the predicate and attaches to its result state. In the case of a reversal of direction, however, such a result state does not necessarily exist. A verb like "write" in (2) is not a lexical accomplishment predicate and does not a have a result as "open" in ((1)) does. Thus, the second reading that "write" used to have cannot be accounted for by the structural approach. We can conclude that there used to be a second lexical entry for counterdirectional *wieder* and *again* and that these have vanished from the lexicon. Note, however, that the participants still understand the sentences' meaning and do not rate the sentences as bad as the ungrammatical fillers. All in all, none of the presented semantic theories is able to explain the results of the study by itself. A hybrid approach seems more promising and able to grasp the diachronic development of *wieder* and *again*.

References

1. Beck, S.: There and Back Again: A Semantic Analysis. Journal of Semantics. 22, 3–51 (2005)
2. Beck, S., Berezovskaya P., Pflugfelder, K.: The Use of Again in 19th-Century English versus Present-Day English. Syntax. 12, 193–214 (2009)
3. Fabricius-Hansen, C.: Wi(e)der and Again(st). Audiatur Vox Sapientiae. In: Fery, C., Sternefeld, W. A Festschrift for Arnim von Stechow. 101–130 (2001)
4. Gergel, R., Beck, S.: Early Modern English again - A corpus study and semantic analysis. English Language and Linguistics. (2012)
5. McCawley, J.: The Role of Semantics in Grammar. Universals in Linguistic Theory. 124–169 (1968)
6. Rapp, I., Stechow, A. von: Fast "almost" and the visibility parameter for functional adverbs. Jorunal of Semantics. 16, 49–204 (1999)
7. Stechow, A. von: The different readings of wieder 'again': A structural account. Journal of Semantics. 13, 87–138 (1998)

Counterfactual Reasoning*

Benjamin Sparkes

Institute for Logic, Language and Computation, University of Amsterdam

1 Introduction

The difficulties in explaining which properties of the world are relevant when making a counterfactual assumption, and more generally evaluating counterfactual conditionals, are well known. I open the paper with a variation on the Jones scenario from [11] that illustrates the complexity of this task, and which serves a counterexample to a number of well-known theories, such as Kratzer [5] and Lewis [7].[1] The theory of Veltman [11] is a notable exception. In this paper I present a simple counterexample to this theory through a small change to the scenario. Therefore an alternative analysis is called for.

This paper develops an epistemological analysis of counterfactuals, and their semantic underpinnings are left aside.

Hence, our focus is on *acceptability*, and not truth, conditions of counterfactuals.[2] Still, the theory I propose has a significant semantic parallel. For, it builds on the basic idea of premise semantics [5,8,11] by determining a 'premise set' from which maximal subsets consistent with the antecedent of a counterfactual can be derived, and against which the consequent can be tested. However, and notably, a comprehensive analysis of counterfactuals is given independent from semantic concerns.

2 Jones' Hat

Consider the following scenario, modified from [11]:

> Each morning Jones wakes up and opens the curtains to see if the weather is good or bad and then flips a coin. Jones is possessed of three dispositions as regards wearing his hat:
> (1) If the weather is fine and the coin lands tails, Jones does not wear his hat.
> (2) If the weather is fine and the coin lands heads, Jones does wear his hat.
> (3) If the weather is bad, Jones wears his hat (regardless the result of the coin toss).
>
> Today the weather is bad, heads came up, and Jones is wearing his hat.

* Thanks to Frank Veltman and Maša Močnik for helpful comments and discussion.

[1] Concerning the latter, see [11, Sec. 2]. The former is discussed below in §3.

[2] For an insightful discussion of reducing semantic problems to epistemology see [3].

Jones' current status corresponds to his third disposition, but do you accept the sentence '*if the weather had been fine, Jones would have been wearing his hat*'? I believe the answer will be yes, though I do not think assent will have been given without hesitation.

The state of the weather both epistemically and ontologically precedes the result of the coin toss. The counterfactual antecedent, then, '*breaks*' from the actual world before the toss lands and Jones' disposition is determined.[3] Assent to the sentence shows that particular matter of fact can have an important role in one's assessment of counterfactuals, even if that fact is temporally posterior to, and causally independent from, the counterfactual antecedents break.[4] Hence, it seems that temporal and causal notions cannot play a sufficient explanatory role. For this reason the relation between *laws*, such as Jones' dispositions, and *facts*, such as the outcome of the coin toss, are our primary concern.

2.1 Facts and Laws Introduced

The author ascribes to the view that counterfactuals have an 'ontic' reading; one considers only the (known) structure the world, and not those beliefs one holds of it. For this reason the assessment of counterfactuals faces problems similar to other sentential constructions that use modal verbs, such as *might* under their ontic reading.[5] For, one cannot objectively establish the properties one ascribes to a situation. Oswald *did* shoot Kennedy, and Kangaroos *do* have tails and so we cannot empirically investigate the contrary. Even if some aspect of the world guarantees the acceptability, or indeed truth, of a counterfactual, the reason for its acceptability, or truth, is (generally) not open to empirical inquiry. There is, then, a tension in counterfactual reasoning. For, one wishes to establish whether a counterfactual conditional holds according to fact, but without epistemic access to any adequate factual base. It is for this reason that laws are important; they express those aspects of the world that are immutable when assessing a counterfactual – typically ruling out *logically* possible situations on account of empirical connections between distinct propositions. In the above scenario these are Jones' dispositions. In other cases they are generalisations of a given regularity (natural laws), matters established by convention (the rules of chess), or principles one adheres to (such as the presumption of rationality in economic reasoning). Regardless the status of a law, then, the sentence expressing it must be true under any counterfactual possibility.

With laws taken to be immutable, the tension noted above follows from determining the set of facts. For, with this set, one can say that a counterfactual is true if its consequent holds in every world where its antecedent, along with the laws and an appropriate subset of facts true of the actual world, holds.

[3] See [1] for a detailed discussion of scenarios such as the one presented.

[4] Note, switching the order of Jones' actions presents firmer intuitions, cf. [11, p. 164].

[5] Consider, for example, the sentence: '*it might be raining*'. The epistemic reading considers whether one has reason to believe it is raining, such as a weather report. The ontic reading, whether the meteorological facts are such that it is possible for it to be raining. The two readings will not necessarily have the same truth value.

Goodman pioneered this approach in attempting to determine the facts 'cotenable' with the counterfactual antecedent [4], in contrast, e.g. to the similarity approach in the tradition of Stalnaker-Lewis, cf. [1, Secs. 3–4]. However, Goodman's solution was circular, and though many theories have been developed in the wake of Goodman's attempt, most (if not all) fail to deal with scenarios such as that of Jones, above. As mentioned above, an important representative, Kratzer [5], will be discussed below. The theory of Veltman [11] is notable for dealing with a wide range of scenarios, including the above. However, Veltman's theory is open to a simple counterexample. Consider the following scenario:

> Each morning Jones wakes up and opens the curtains to see if the weather is bad or fine and then flips a coin. Jones is possessed only of a single disposition as regards wearing his hat: If the weather is fine and the coin lands tails, Jones does not wear his hat.
>
> Today the weather is bad, tails came up, and Jones is wearing his hat.

This scenario is as before, but such that Jones is only possessed of his first disposition, and the result of the coin toss is tails. Given this scenario, Veltman's theory predicts one should accept the counterfactual statement: *if the weather had been fine, Jones would have worn his hat if and only if the coin toss had not landed tails*'.[6] This should not be the case. It is easy to verify that Jones' second and third dispositions are irrelevant to the scenario, and so '*if the weather had been fine, Jones would not have been wearing his hat*' is supported.

2.2 The Epistemic Approach

Given the shortcomings of the above theories, then, an alternative is called for. Furthermore, we note that each of the theories has a distinct epistemological aspect. Goodman relies on the notion of 'projectiblility', cf. [4, Ch. IV]. Kratzer appeals to cognitive processes in a number of details, notably requiring the cognitive viability of 'base sets', which are fundamental to her analysis of counterfactuals [5, p. 133]. The Stalnaker-Lewis approach requires an account of comparative similarity, relying on (in some cases) a system of weights and measures, e.g. [7]. While Veltman relies on the epistemic process of 'retraction' [11, pp. 168–169].

Our approach takes this epistemic aspect as its interest, and, in particular, the role of inferential rules. For our purposes we characterise which sentences are laws by the epistemic attitude one takes toward them; sentences that expresses some aspect of the world which one is not prepared to give up – i.e. consider false – when making a counterfactual assumption.[7] Similarly, facts are taken to be the *known* facts. Because of the nature of the approach taken, the argument presented below gives a clue to, but does not directly deal with, those aspects of language one must understand to give an adequate account of the semantic and/or pragmatic content of counterfactuals.

---

[6] For those familiar with the framework of Veltman [11], using the shorthand defined below: $\mathbf{1}[\Box((f \wedge t) \rightarrow \neg w)][\neg f][t][\textit{if it had been} f] \vDash ((\neg t \wedge w) \vee (t \wedge \neg w))$.

[7] We consider assessing counterfactuals to be a processes similar to the Ramsey test for conditionals [10, p. 106]. See §4.4 for our implementation.

2.3 Notation

Before continuing we introduce some notational shorthand. Let f be short for *'The weather is fine'*, t for *'The coin lands tails'*, and w for *'Jones wears his hat'*. We use '$\Box$' to shorten *'it is a law that ...'*. Hence, $\Box(f \to (t \to w))$ reads *'if the weather is fine, then if the coin lands tails then Jones wears his hat'*, and Jones' dispositions can be formalised likewise. $\phi \rightsquigarrow \psi$ abbreviates '*if it had been the case that* ϕ *it would have been the case that* ψ*'*.

Our formalisation syntactically follows the propositional calculus, yet, as will become clear below, we only require the connectives obey certain rules of inference (these will be detailed in §4). Hence, '$\to$' need not be understood as material implication, nor '$\vee$' as (inclusive) disjunction, etc., though we shall (implicitly) take their customary interpretation. Put another way, the propositional calculus is syntactically, if not semantically, rich enough to support our analysis of counterfactuals. Hence, as we do not require a semantic component we accept, but do not commit ourselves to, that of the calculus.

3 Facts and Laws Analysed

Fundamental to our analysis is the observation that distinct, deductively equivalent, formulations of laws structure situations in different ways. To illustrate, consider again Jones' second disposition. This is formalised as $f \to (\neg t \to w)$ and is (deductively) equivalent to $(f \to w) \vee (\neg t \to w)$. Yet, one understands Jones' dispositions such that both the weather and the coin must be settled to determine whether he wears his hat. However, these sentences will be true in the same set of possible worlds, and hence express the same proposition. Similarly, one is unwilling to accept $f \to (\neg w \to t)$ as descriptive of Jones' disposition. For, intuitively, it reads, on the condition the weather is fine, Jones' decision not to wear his hat determines that the coin toss lands tails. If these sentences do structure situations in distinct ways, then they must be understood differently.[8] How, then, can distinct sentences that express the same proposition in turn express distinct aspects of the world?

This observation is not new, and is fundamental to the approach of Kratzer [5] and Rescher [8]. Kratzer observes that inductively established laws, such as 'all ravens are black' and 'all non-black things are non-ravens' are confirmed by distinct observations, but are logically equivalent, and can both be falsified by equivalent sentences. Kratzer's appeal to inductive logic echoes that of Rescher. However, Rescher's note that $P(X/Y) = P(\overline{Y}/\overline{X})$ is not a theorem of probability theory [8, p. 80] cuts cleaner than Kratzer's arguments. Regardless, both Kratzer [5, Sect. 5.5] and Rescher [8, Appx. 3.4] take an observation regarding inductive reasoning to bear the explanatory burden between distinct sentential formulations of laws. While it has long been recognised that laws cannot be

[8] Furthermore, there should, intuitively, be a corresponding semantic difference between them. However, our focus is on acceptability and not truth.

formulated using universally quantified material conditionals, Kratzer's explanation is modelled on these and takes information about 'confirmation sets' to be embedded in the semantics [5, Sec. 5.5]. Regardless of whether Kratzer's account is semantically adequate, and whether confirmation sets are encoded into the meaning of sentences, it is puzzling why these problems related to inductive reasoning should play a prominent role in the deductive exercise of reasoning about counterfactuals. The sentences taken to be laws are often not descriptions of a given regularity, but fixed by context (as noted above). Our task is to understand why distinct sentences that express the same propositions in turn express distinct aspects of a situation, but appealing to induction limits us to laws of the natural world. Philosophical objections aside, Kratzer's proposal may be reinterpreted under a non-inductive guise. However, it is questionable if confirmation sets are an adequate formal base for use in analysing counterfactual reasoning. With Jones' dispositions one wishes to distinguish between f, t and w. For, it is only the first two facts that actively structure the situation, and w must be true given these. Yet, all must be present in a confirmation set. Therefore, one cannot distinguish t to be preserved under the counterfactual assumption of $\neg f$. Indeed, Kratzer's naïve theory [5, p. 133] predicts $\neg f \rightsquigarrow w$, while Kratzer's refinement [5, p. 151] predicts $\neg f \rightsquigarrow ((t \wedge \neg w) \vee (\neg t \wedge w))$. It is, then, unclear whether confirmation sets can yield sufficient semantic structure to analyse counterfactuals.

4 Analysis

Let us now turn to providing an answer to the question, set above, of how distinct sentences which express the same proposition can in turn express distinct aspects of the world. Our analysis does not look so far afield as inductive logic, but relies on the observation that distinct sentential formulations of a proposition can behave in (remarkably) divergent way when subjected to a restricted class of deductive inference rules. In turn we will understand the importance of the laws and how one reasons with them, and then which facts one keeps under a counterfactual assumption.

Here lies the heart of our approach: given a set of sentences that are taken to express laws, $\mathcal{L}$, one can determine two subordinate sets of sentences. First, those sentences that are immediate consequences of the laws and second, those sentences that, structure the world in accordance with the laws. For example, $\Box(\phi \wedge \psi)$ determines both ϕ and ψ must be true. While, if $\Box(\phi \rightarrow \psi)$ and ϕ is true, then ψ must also be true. Below we shall combine these two sets to one, $\mathcal{A}$, the 'active' facts relative to $\mathcal{L}$. The motivating thought is that laws mark certain facts as (potentially) active in a situation, while other facts are inactive. Those active facts, if true, bring about other facts, while inactive facts can not. The facts that one wishes to keep hold of when making a counterfactual assumption will be derived from these 'active' facts.

This can be illustrated by, again, considering Jones' dispositions from the opening scenario. Fine weather ensures that Jones will premise his decision to

wear his hat on the result of the coin toss, while bad weather ensures Jones will wear his hat regardless. Hence both f and $\neg f$ 'actively' structure the world, according to Jones' idiosyncrasies. The former ensures both $t \rightarrow \neg w$ and $\neg t \rightarrow w$ hold, and the latter that w holds. In turn, given f, both t and $\neg t$ structure the world to be such that either w or $\neg w$.

4.1 Decomposing Laws

From an inferential point of view there are only so many ways one can decompose information contained in a law taken as a sentence. For example, one may only apply those inferential rules such that the conclusion is no more complex than any premise. Indeed, in the Jones scenario the facts $f, \neg f, t$, and $\neg t$ are just those from which aspects of the laws follow under the inference rule of *modus ponens*. Jones' dispositions taken as laws, and the rule of *modus ponens* in hand, can reduce Jones' dispositions to their most basic parts with certain (combinations of) facts. However, neither w nor $\neg w$ are such facts, for both only occur as the consequent of a (conditional) law. These are inactive for they are each determined *after* the arrangement of $f, \neg f, h$, and $\neg h$ has been settled and so cannot be used in analogous reductions. For this reason, we consider $f, \neg f, t$, and $\neg t$ to be the active facts, given Jones' dispositions. Reasoning about the structure of a situation, given some set of laws, is, under this view, inherently a reductive activity. While *modus ponens* is the fundamental inference rule used in Jones' scenario our interest, more broadly stated, *is* in those inferential rules such that the conclusion is no more complex than any premise. For, those active facts identified will then in conjunction with the laws entail other facts, specifically facts whose relation to other facts is described by the laws. These are exactly those facts we expressed an interest in above. Furthermore, laws, under this restriction, cannot be manipulated into truth functionally equivalent statements, and moreover one's reasoning is restricted to only those facts described in the law – for example, the introduction of new facts by certain forms of reasoning such as *reductio ad absurdum* cannot occur.

4.2 Polarity

Yet we have not considered rules such as *modus tollens*. This allows one to derive t from f and $\neg w$ in a manner similar to the inadequate description of Jones' disposition as $f \rightarrow (\neg w \rightarrow t)$, above. Perhaps there is some property of *modus tollens* that excludes it as an applicable rule of inference when decomposing certain uses of conditionals. Our proposal is that the *polarity* of a fact matters. This distinction relies on taking a and $\neg a$ to express information concerning the same fact under different polarities, the former *positive*, and the latter *negative* polarity. More generally a sentence can either be positive or negative, but we relate any formula with it's negation in this manner. Our concern below will be how this distinction relates to facts, and so we will keep the more restricted terminology It matters, then, that any derivation of a fact must occur in the conclusion with the same polarity as it has in its occurrence as part of a law.

Clearly it is an arbitrary choice whether one represents a fact with or without a negation sign. For example, consider 'the coin lands tails'. Above (§2.3) we represented this with the formula t, and hence represented 'the coin lands heads' with $\neg t$. However, we could have represented the latter with h and so the former with $\neg h$. We could also have formalised the former by t, the later by h and related these two formulas by the following (classically equivalent) laws: $h \leftrightarrow \neg t$ and $t \leftrightarrow \neg h$. Issues such as these confuse the issue at hand. Our observation is this: $\neg\phi$ is distinguished from ϕ, and if a law concerns ϕ then it does not (ipso facto) concern $\neg\phi$ although one can find a formula that is deductively equivalent and contains $\neg\phi$ given the full calculus. The particular polarity of a fact matters only in the relation between facts and laws. When analysing a scenario the particular polarity chosen in the formalisation can be made arbitrarily, so long as this choice is consistent. Hence, we argue laws only carry (non-inferential) information concerning the polarity of a fact as it occurs in the law, and *modus tollens* does not preserve polarity – for it allows one to infer $\neg\phi$ from $\phi \rightarrow \psi$ and $\neg\psi$, but $\neg\phi$ and ϕ differ in polarity. Attention to polarity explains why Jones' disposition cannot be reformulated, and why *modus tollens* cannot be applied as a rule of inference. And in turn, the importance of polarity is certainly justified by inductive considerations, as argued for by Kratzer and Rescher[9] but also by other uses of conditionals.

Consider as an example the conditional command: '*if you see smoke, shout fire*'. Silence is suggested in the absence of smoke. The command is not equivalent to '*if you do not shout fire, there is no smoke*', even though in the absence of smoke, truth functionally, both shouting fire and remaining silent are permissible. Hence, we observe a similar phenomena to that of Jones' dispositions. A common observation is that conditional sentences often invite one to treat implication as, at first glance, bi-implication (cf. [2]).[10] Hence, one supports the command above with '*if there is no smoke, I do not shout fire*'. The explanatory adequacy of this approach is questionable in this case, because one does not seem to reason from the command to the conclusion that *if I do not see smoke, I do not shout fire*. Rather, because shouting fire is premised on seeing smoke, one has no information regarding the constraints on one's actions in smokeless situations. Hence, in such situations one has no reason to *infer* they should not shout fire, not that one thinks it permissible to infer they should not shout fire.

Polarity of a fact matters. In counterfactual reasoning, one needs justification, to keep hold of the facts that matter, just as one needs to sight fire for the conditional to take effect. Thus, polarity affects the set of facts that one is

[9] Cf. Rescher's note, observed above in §3 – the probability of X conditional on Y does not (in general) carry (exhaustive) information about the conditional probability of *not-Y* given *not-X*.

[10] To be sure, the same explanation may be tentatively raised for Jones' dispositions. It seems to me that this approach is ultimately unsuccessful. However, aside from the comments made above, which dispute this approach in relation to conditional commands, I will not discuss this topic further.

justified in keeping hold of under a counterfactual assumption, for polarity affects whether a fact is active (with respect to a set of laws) or not.[11]

4.3 A Class of Inferential Rules

The class of inferential rules one can use to determine the active facts are, then, those such that the conclusion is no more complex than any premise, in which the conclusion is contained in the premises, and such that the polarity between the occurrence of a sentence in the conclusion is equal to its occurrence in the premises. Common examples of such rules are *modus ponens, simplification, disjunction elimination, disjunctive syllogism, constructive dilemma, hypothetical syllogism,* and *biconditional elimination.*[12] These rules allow one to exploit and preserve the information present in the structure of a sentence. Furthermore, they preserve the sentential form of the laws and facts one takes as premises. This has been seen to be essential. Each such rule of inference allows one (if possible) to derive some aspect of a law, some fact they law describes relative to other facts, with the same polarity as it occurs in the law. There is not adequate space to attempt a rigorous justification of the foregoing selection beyond what has been said above. I leave it for further investigation to determine the exact set of inference rules. Yet, each of the foregoing can be found requisite for certain scenarios and given intuitive justification on a case by case basis.

Finally, we observe the epistemology (or semantics) of inductive reasoning may be a special case of the semantics which grounds the above approach. For example, it will be the case that for every inductive law, the active facts will be elements of the confirmation sets used by Kratzer.

4.4 Definitions

Let us now establish a precise formulation of the foregoing ideas. The following two definitions lay the groundwork:

Definition 1: Worlds, laws, facts. Let $\mathcal{S}$ be a finite set of atomic sentences. The set of formulas is the closure of $\mathcal{S}$ under the connectives $\neg, \wedge, \vee, \rightarrow$, and $\leftrightarrow$. A *world* is a maximally consistent set of formulas.[13] The laws and facts form consistent (possibly empty) sets of formulas. Let $\mathcal{U}$ denote the set of worlds, $\mathcal{L}$ the set of laws, and $\mathcal{F}$ the set of facts. We denote by $\mathcal{U}^{\mathcal{L}}$ the set of worlds in which the laws hold: $\{w \in \mathcal{U} \mid \forall \ell \in \mathcal{L}, \ell \in (w \cap \mathcal{L})\}$. $\dashv$

[11] Furthermore, though Jones' dispositions invite a causal, or temporal, structuring the conditional command is not descriptive, and so such notions do not apply, at least not in the way they would apply to Jones' dispositions. We have shown, however, that both may share a common explanation and therefore, there seems little need to appeal to such notions.

[12] The definitions introduced below (in particular the implications of footnote 15) place an implicit restriction on this class, such that whimsical rules, for example strengthening *modus ponens* with arbitrary many irrelevant antecedents, are irrelevant.

[13] A set of formulas Γ is consistent if $\neg\exists\phi[\Gamma \vdash \phi \wedge \neg\phi]$ and inconsistent otherwise, maximally so if $\forall\Gamma'[\Gamma \subset \Gamma']$, Γ' is inconsistent.

Definition 2: Consequence relations. The following allow us to state the class of inference rules applied to sets of formulas. Let Φ be a set of formulas.

(1) $\Phi \vdash \phi$ if ϕ can be inferred from Φ by the application of any rule of inference.
(2) $\Phi \vdash_e \phi$ if ϕ can be inferred from Φ by the rules allowed in §4.3.[14] ⊣

The following quintet of definitions show how information about the structure of facts can be derived, and how the set, $\mathcal{A}$, of 'active' facts can be established.

Definition 3: Activity. Let ϕ and ψ be arbitrary formulas and Ψ, Φ sets of formulas consistent with $\mathcal{L}$, the set of formulas taken as laws.[15] The closure of Φ under subformulas is denoted by $c(\Phi)$.

(1) Ψ *establishes* ϕ if $\mathcal{L} \cup \Psi \vdash_e \phi$, $\phi \notin \Psi$, and $\neg\exists \Phi \subset \Psi$ such that $\Phi \vdash_e \phi$.
(2) The *basis* of ϕ, $b(\phi)$, is the set $\bigcup\{\Psi \mid \Psi \text{ establishes } \phi\}$.
(3) The *scope* of ϕ is the closure of its basis under subformulas and (single) negation $c(b(\phi))$.
(4) ψ is *active* with respect to ϕ if $\psi \in c(b(\phi))$.
(5) ψ is *active with respect to* Φ if $c(b(\psi)) = \emptyset$ and $\exists \phi \in \Phi : (\psi \in c(b(\phi)))$. ⊣

Each definition calls for remark: Definition 3.1 rests on the idea presented in § 4 above, that one can determine the set of active facts from those that can be used as premises when restricted to certain inference rules. In the terminology used above, each element of Ψ is 'active' in establishing ϕ, and conditions for non-circularity are added. In general there will not be a unique such Ψ. Hence, 3.2 collects all such ways of establishing ϕ together. 3.4, activity understood here in relation to counterfactual reasoning, does not distinguish the polarity of a fact. That distinction was made in §4.2, here we are not concerned with which facts are described by the laws but with which facts establish other facts. When reasoning counterfactually one wishes to preserve the truth of those facts that *can* affect the structure of the world according to the laws. 3.3 and 3.4 allow one to do this. We determine those facts that, if true (or false), make a difference with respect to a formula. 3.5 generalises 3.4 to sets of formulas, and ensures no active fact depends on the truth of another fact, we return to this definition in §6 this ensures our notion of activity as defined applies only to literals.

Definition 4: Counterfactual Reasoning. When reasoning about counterfactual worlds what one considers possible is not only constrained by the laws one takes to hold, but also the active facts with respect to the actual world. Let $\mathcal{A}^{\mathcal{L}}$ denote the set of active formulas relative to $\mathcal{L}$, by definition 3.5.

(1) $\mathcal{A}^{\mathcal{L}}_{\mathcal{F}} = \mathcal{A}^{\mathcal{L}} \cap \mathcal{F}$, the set of formulas active in the actual world.
(2) $\mathcal{U}^{\mathcal{L}}(\Phi) = \{w \in \mathcal{U}^{\mathcal{L}} \mid \Phi \subseteq w\}$, the set of worlds such that Φ is a subset.

[14] One may also admit certain rules of replacement, e.g. associativity and commutativity. Undischarged assumptions are not permitted.

[15] Strictly speaking, we require Φ and Ψ to be *situations*, effectively 'small worlds'. The set of situations is $\{s \mid \exists w \in \mathcal{U}^{\mathcal{L}} : s \subseteq w \wedge \neg\exists s'(s' \subseteq w \wedge s' \subseteq c(s) \wedge s \subset s')\}$.

(3) $\mathcal{U}^{\mathcal{L}}_{\mathcal{F}}(\Phi) = \{w \in \mathcal{U}^{\mathcal{L}}(\Phi) \mid \neg\exists w' \neq w(w' \in \mathcal{U}^{\mathcal{L}}(\Phi) \wedge ((w \cap \mathcal{A}^{\mathcal{L}}_{\mathcal{F}}) \subset (w' \cap \mathcal{A}^{\mathcal{L}}_{\mathcal{F}})))\}$.
(4) '*if it had been* ϕ' $= \mathcal{U}^{\mathcal{L}}_{\mathcal{F}}(\{\phi\})$, the set of relevant worlds to a counterfactual assumption ϕ.
(5) $\phi \rightsquigarrow \psi$ iff $\forall w \in \mathcal{U}^{\mathcal{L}}_{\mathcal{F}}(\{\phi\}), \psi \in w$, the counterfactual conditional. ⊣

5 Applications

5.1 Jones' Hat

Following the shorthand defined above the set of laws in the opening scenario is $\{(f \wedge t) \rightarrow \neg w, (f \wedge \neg t) \rightarrow w, \neg f \rightarrow w\}$, and the set of facts is $\{\neg f, \neg t, w\}$. By definition 3.1, the set $\{f \wedge t\}$ establishes $\neg w$, and the sets $\{f \wedge \neg t\}$ and $\{\neg f\}$ establish w, and no facts other than w and $\neg w$ are established in this sense. Hence, by 3.2 $\{f \wedge t\}$ is the basis of $\neg w$ and $\{f \wedge \neg t, \neg t\}$ is the basis of w. By 3.3 the scope of these bases is their closure under subformulas, and so we obtain the sets, omitting conjunctions for brevity, $\{\neg(f \wedge t), f \wedge t, f, t, \neg f, \neg t\}$ and $\{\neg(f \wedge \neg t), f \wedge \neg t, \neg t, t, \neg f, f\}$ respectively. By 4.1, the formulas active in the actual world is the set $\{\neg f, \neg t\}$, as the sets of active formulas relative to the laws of the scenario is $\{f, t, \neg f, \neg t\}$. By 4.4 the counterfactual assumption '*if the weather had been fine*' is $\mathcal{U}^{\mathcal{L}}_{\mathcal{F}}(\{f\})$. This, by 4.3 is just the set of worlds containing as many formulas active in the actual world as possible given the counterfactual assumption.

We know, then, that $\{f, \neg t\}$ must be a subset of each (counterfactually relevant) world. This means that the (unique) counterfactual world is $\{f, \neg t, w\}$, following Jones' second disposition. Therefore, '*if the weather had been fine, Jones would have been wearing his hat*'.

5.2 King Ludwig and the Three Sisters

With respect to the claims of this paper the following two scenarios are of particular interest, termed below 'King Ludwig' and 'the Three Sisters' respectively.

> King Ludwig of Bavaria likes to spend his weekends at Leoni Castle. Whenever the Royal Bavarian flag is up and the lights are on, the King is in the Castle. At the moment the lights are on, the flag is down, and the King is away. Suppose now counterfactually that the flag were up. [5, p. 140]

> Consider the case of three sisters who own just one bed, large enough for two of them but too small for all three. Every night at least one of them has to sleep on the floor. At the moment Billie is sleeping in bed, Ann is sleeping on the floor, and Carol is sleeping in bed. Suppose now counterfactually that Ann had been in bed ... [11, p. 178]

As Veltman notes the two scenarios share the same logical structure. We identify the law in the former scenario as '*if the Royal Bavarian flag is up and the lights are on, the King is in the Castle*' and in the latter as '*at least one*

of the sisters must sleep on the floor' – when modelled (naturally) using the propositional calculus – both express a propositionally equivalent sentence. For example, the latter can be reformulated as '*if Ann sleep and Billie sleep in the bed, Carol sleeps on the floor*', demonstrating the underlying equivalence.

However, the first two laws differ in the facts they deem active, while the first and third laws mentioned are equivalent in this respect. Taking the first two laws we predict that '*if the flag were up then the King would be in the castle and the lights would still be on*' and '*if Ann had been in bed, either Billie or Carol would have been sleeping on the floor*'. While, if the reformulated sentence is taken as the (singular) law in the three Sisters scenario we predict '*if Ann had been in bed, Carol would be sleeping on the floor*'. To be sure, reformulating the scenario such that the third law is the only constraint mentioned supports this counterfactual assessment, and the previous two predictions follow intuition. We have, then, a clear example of the importance of distinguishing distinct sentential formulations of the same proposition.

5.3 Further Scenarios

To conclude this section, we note the ambiguity observed in Lewis' barometer scenario [6, pp. 564–565] can be explained by whether one takes the conditional '$air\ pressure \rightarrow reading$' or '$reading \rightarrow air\ pressure$' to express the relevant law.

Furthermore, the theory offers the same prediction as Veltman's for his Duchess scenario [11, p. 174] but does not make contrary to intuition predictions as his theory does with Schulz's circuit example [9, p. 244], and deals with many other cases found in the literature. However, the theory presented represents a fragment of what is required to give a comprehensive account.

6 Further Research and Conclusion

We raise three issues for further research.

First, do laws that fall under the scope of a negation structure facts? If so, the theory requires refinement. However, note there is no restriction on taking any number of distinct, but propositionally equivalent, sentences to express laws. I have argued only that, in full generality, the content of a law is not preserved under propositional equivalence. Hence, there is no restriction, nor *a priori* reason, for not taking $\neg(a \wedge b \wedge c)$ and $\neg a \vee \neg b \vee \neg c$ to both express laws.

Second, definition 3 ensures that only literals will be active facts, and it is questionable if this restriction is well motivated. One may argue the constraints of definition 3.5 should be weakened to those of definition 3.4, generalised to sets of formulas. Hence, if definition 3.4 is taken as our notion of activity, and active facts are permitted to be established by other facts. If this were the case then given $\Box(a \rightarrow (b \rightarrow (c \rightarrow d)))$, the active facts are $a, b, c, b \rightarrow (c \rightarrow d), c \rightarrow d$, and their negations. Let a, b, c, d hold in the actual world. Reflection shows, counterfactually assuming $\neg b$, d *may* be the case, but not that it *would* be the case. This is because b is only a sufficient condition, and not necessary, for d

to be the case. Hence, counterfactually assuming $\neg b$ leaves us without reason to think d *will* remain the case. Yet, as c and $c \rightarrow d$ would then be active in the actual world (both partly establish d), the system would predict $\neg b \rightsquigarrow d$. Issues such as these may be (partially) accounted for by dispreferring worlds in which implication 'trivially' hold because of the truth of the consequent, or giving preference to worlds in which certain, or a greater number, of complex sentences obtained from the laws hold. More generally one may look to worlds that 'conform' to the laws that hold in an actual world, in a yet to be specified sense. Yet, the most natural assumption is to restrict the notion of activity to literals. Careful investigation is required, but the epistemological, and syntactic, approach advocated in this paper has much to promise in this regard.

Finally, tautologies can not (in general) be taken as laws, given the above definitions. For example, it is easy to see that adding the premise that $(\neg w \wedge \neg w) \rightarrow \neg w$ to Jones' scenario would void our predictions, by rendering $\neg w$ an active fact. Tentatively one may argue that one does not take as laws sentences which hold by logical necessity, for these cannot affect the structure of the world. However, this is an observation that requires further attention, even if a natural explanation can be given.

The core idea of this paper is that significant insight into counterfactuals is gained by understanding how people reason with the information they have – the facts they know, and the laws they take to hold. We conclude by noting that, following the ideas presented, the attitudes one takes towards the world by whether, and which, sentences are taken as laws, and which facts are known will constitute a significant source of vagueness and context sensitivity for counterfactuals.

References

1. D. Edgington. Counterfactuals and the Benefit of Hindsight. In *Cause and Chance: Causation in an Indeterministic World*. Routledge. (2004)
2. M. L. Geis and A. M. Zwicky. On invited inferences. *Linguistic inquiry*, 2(4):561–566. (1971)
3. A. Gillies. Epistemic Conditionals and Conditional Epistemics. *Noûs*, 38(4):585–616. (2004)
4. N. Goodman. *Fact, Fiction & Forecast*. University of London, London. (1954)
5. A. Kratzer. An Investigation of the Lumps of Thought. In *Modals and Conditionals: New and Revised Perspectives*. OUP, Oxford. (2012)
6. D. Lewis. Causation. *Journal of Philosophy*, 70(17):556–567. (1973)
7. D. Lewis. Counterfactual Dependence and Time's Arrow. *Noûs*, 13(4):455–476. (1979)
8. N. Rescher. *Hypothetical Reasoning*. Amsterdam, North-Holland Pub. Co. (1964)
9. K. Schulz. "If you'd wiggled A, then B would've changed". *Synthese*, 179(2):239–251. (2011)
10. R. C. Stalnaker. A theory of conditionals. In *Studies in Logical Theory*, pages 98–112. Blackwell. (1968)
11. F. Veltman. Making Counterfactual Assumptions. *Journal of Semantics*, 22(2):159–180. (2005)

A Proof-Theoretic Approach to Generalized Quantifiers in Dependent Type Semantics*

Ribeka Tanaka

Ochanomizu University†
tanaka.ribeka@is.ocha.ac.jp

Abstract This paper presents a formalization of generalized quantifiers (GQs) in the framework of dependent type semantics (DTS), a proof-theoretic semantics for natural language. DTS is a system that extends dependent type theory with a dynamic context-passing mechanism. In this study, we give an appropriate and simplified semantic representation for the determiner *most*, which is a crucial example of GQs, and for other determiners including numerical determiners. We also prove that the determiner *most* satisfies conservativity in our dynamic setting.

1 Introduction

Generalized quantifiers (GQs) are well studied within the model-theoretic approach to natural language semantics. In this approach, the meanings of GQs are defined as relations between sets, and various logical properties of determiners such as conservativity and monotonicity have been established [3]. This model-theoretic conception of GQs was also adopted by discourse representation theory [12] and applied to discourse phenomena such as donkey anaphora. In contrast to the model-theoretic approach, there is a proof-theoretic approach to natural language semantics ([18],[17],[13],[5]). The proof-theoretic approaches are attractive in that entailment relations can be defined directly without appealing to models. However, as compared with model-theoretic analyses, the comprehensive analysis of GQs within proof-theoretic semantics that can account for various linguistic phenomena is still underdeveloped.

There are two kinds of approaches to treating GQs in a proof-theoretic framework. One is to construct a proof system that contains determiners as primitives; the study of natural logic (e.g., [6]) can be subsumed under this approach. Currently, however, such a proof system is not concerned with dynamic linguistic phenomena such as anaphora and presupposition.[1] Another approach is to give

* I would like to express my gratitude to Daisuke Bekki, Koji Mineshima and Pascual Martínez-Gómez for helpful discussions and supports all along the writing of this paper. I am grateful to the three anonymous reviewers for their detailed and insightful comments and suggestions. This research was supported by JST, CREST.

† Graduate School of Humanities and Sciences, Faculty of Science, 2-1-1 Ohtsuka, Bunkyo-ku, Tokyo 112-8610, Japan.

[1] Recently, Francez and Ben-avi [9] proposed a natural deduction system for GQs and proved logical properties of determiners including conservativity. However, to explain dynamic phenomena was not within the scope of their study.

an explicit definition of determiners. In particular, Sundholm [19] presented such a definition within the framework of constructive type theory [14,15], a framework that has been successfully applied to the dynamic aspects of natural languages ([18],[17],[4,5]). Note that even if a determiner has a fixed model-theoretic meaning, to give an explicit definition for it within a type-theoretic framework is not a trivial task; in particular, it is challenging to give semantic representations that maintain both logical and linguistic properties of determiners to account for their dynamic behavior. In this paper, we adopt this second explicit approach and argue that it is a viable one to natural language semantics.

Sundholm [19] sketched the constructive definition of *most* in a way that can avoid the so-called proportion problem [11]. Tanaka et al. [20] improved Sundholm's analysis in several respects and gave a semantic representation of *most* in the framework of dependent type semantics (DTS), which is an extension of dependent type theory (DTT) in terms of natural language dynamics. However, there remain several issues in their analysis. First, it makes an undesirable prediction for plural anaphoric references to NPs with determiner *most*. Second, Tanaka et al. [20] does not deal with determiners other than *most*. It is a nontrivial assertion that other GQs can be represented in the same manner. Third, they do not discuss logical properties of GQs such as conservativity and monotonicity. In this paper, we therefore address these three issues in Tanaka et al. [20] and give an alternative representation of *most*.

This paper is structured as follows. First, we introduce the formalization presented by Sundholm [19], which is based on DTT. Then we give a brief overview of the framework of DTS and show how dynamics is handled in that framework. We then introduce the formalization by Tanaka et al. [20], and point out its problems. Given this background, we propose a more suitable semantic representation for *most* that captures both weak and strong readings of determiners. We also show that it can be extended to other determiners. Finally, we prove conservativity of *most* in our setting.

2 Constructive Generalized Quantifiers as Considered by Sundholm

Sundholm [19] defines quantifying determiners such as *finitely many A are* φ, *there are at least as many* φ *in A as* ψ *in B* and *most A are* φ in terms of DTT. Noteworthy difference to simply typed theory [2] is that types may depend on terms in DTT. For example, $\texttt{List}(n)$ can be a type of lists of integer-length $n : \texttt{int}$. Therefore, type $\texttt{List}(n)$ is a type that depends on the term n. DTT contains the type constructors Σ and Π. The type constructor Σ corresponds to a generalized form of the product type, and it behaves as an existential quantifier. When $x \notin fv(B)$, $A \wedge B$ is defined as $(\Sigma x : A)B$. The type constructor Π corresponds to a generalized form of the functional type and behaves as a universal quantifier. When $x \notin fv(B)$, $A \to B$ is defined as $(\Pi x : A)B$. $A \leftrightarrow B$ is defined as $(A \to B) \wedge (B \to A)$.

Sundholm [19] defines *most A are* φ (where A represents a noun phrase) as follows:

$$\frac{A : \texttt{set} \quad \varphi : A \to \texttt{Prop} \quad a : \mathrm{Finite}(A)}{\begin{array}{l}\mathrm{Most}(A, \varphi) = (\Sigma k : N)(k \ \geq [\pi_1(a)/2] + 1 \ \& \ (\Sigma f : M(k) \to A) \\ \qquad (\mathrm{injection}(f) \ \& \ (\Pi y : M(k))\varphi(\pi_1(fy)))) : \texttt{Prop}\end{array}}$$

Finite(A) is defined as follows:

$$\frac{A : \texttt{set}}{\mathrm{Finite}(A) = (\Sigma k' : N)(\Sigma f : M(k') \to A)(\mathrm{bijection}(f)) : \texttt{Prop}}$$

Here, the first projection of the term of type Finite(A) (i.e., $\pi_1(a)$) is a natural number corresponding to the cardinality of A. In the definition of $Most(A, \varphi)$, the term $[\pi_1(a)/2]$ indicates the largest natural number less than or equal to $\pi_1(a)/2$; $M(k)$ is a set with cardinality k, where the term k represents the number of least possible majority in A. Mapping f is an injection that maps every element in $M(k)$ to an element in A that satisfies φ. When such a mapping f exists, more than half of the entities in A certainly satisfy φ. This captures the intended meaning of *most*. For detailed definitions, see [1] and [19].

A problem with this definition is that it cannot deal with complex NPs such as *most B who are C*. The restrictor *B who are C* is formalized as $(\Sigma x : B)C$ in the framework of DTT. In this case, *most* needs to count the elements in set B, not the elements (i.e., ordered pairs) in set $(\Sigma x : B)C$. Accordingly, Sundholm [19] introduces the definition of injection relative to a set B, written as B-injection, so as to count only B.[2]

$$\frac{B : \texttt{set} \quad C : B \to \texttt{Prop} \quad D : \texttt{set} \quad f : D \to (\Sigma x : B)C}{B\text{-injection}(f) = (\Pi y : D)(\Pi z : D)(eq(B, \pi_1(fy), \pi_1(fz)) \to eq(D, y, z)) : \texttt{Prop}}$$

In DTT, $eq(T, x, y)$ means that x and y of type T are equal. We can define B-surjection and B-finite in the same manner. This avoids the proportion problem [11], which is often a problematic point in regard to *most* and other determiners. The definition of $Most((\Sigma x : B)C, \varphi)$ is given as follows.[3]

$$\frac{B : \texttt{set} \quad C : B \to \texttt{Prop} \quad \varphi : (\Sigma x : B)C \to \texttt{Prop} \quad a : B\text{-finite}((\Sigma x : B)C)}{\begin{array}{l}\mathrm{Most}((\Sigma x : B)C, \varphi) = (\Sigma k : N)(k \ \geq [\pi_1(a)/2] + 1 \ \& \ (\Sigma f : M(k) \to (\Sigma x : B)C) \\ \qquad (B\text{-injection}(f) \ \& \ (\Pi y : M(k))\varphi(fy)))) : \texttt{Prop}\end{array}}$$

There remains a problem: this definition is based on the assumption that the quantifier *most* counts the elements in the set B with respect to the restrictor represented by $(\Sigma x : B)C$. Accordingly, the same definition is not appropriate when the set to be counted is further embedded in the representation of the restrictor. For instance, consider the restrictor in (1), whose representation is shown in (2).

[2] Since Sundholm's original definition of B-injection contains a type mismatch, we present a slightly modified version of it here.

[3] Sundholm [19] suggested an inductive definition of $Most(A, \varphi)$ for any small type A. The definition presented here is the case in which A is of the form $(\Sigma x : B)C$. As Sundholm himself pointed out, however, this approach requires the case analysis of A, and hence leads to a lack of uniformity in the definition of GQs.

(14) farmers who own a donkey who are rich

(15) $(\Sigma z : (\Sigma x : Farmer)(\Sigma y : Donkey)Own(x, y))Rich(\pi_1 z)$

In this case, a determiner *most* attached to the restrictor (1) needs to count the elements in the set *Farmer* in order to avoid the proportion problem. Schematically, this means that the set B to be counted appears in the form $(\Sigma z : (\Sigma x : B)C)D$. However, it is not clear how to give such an alternative definition of *most* in Sundholm's system. Thus, depending on the form of a restrictor, Sundholm's definition is applied in unintended ways that may cause the proportion problem.

In addition, it is known that donkey sentences have two kinds of reading, *weak* and *strong* readings [7]; but the above definition is only applicable to the weak reading. Furthermore, to account for the interaction of GQs and anaphora, it is desirable to add a mechanism to deal with dynamics of natural language as explained below.

3 Dependent Type Semantics

In this section, we introduce the framework of DTS [4], which is the semantics we adopt for our formalization. The distinctive feature of this system compared with that of Sundholm [19] is that DTS has a context-passing mechanism which enables implementation of dynamics.

DTS is a natural language semantics based on DTT [14,15]. Below, we show some examples of semantic representations in DTS. In DTS, the sentence (3) is represented as in (4):

(16) A man entered.

(17) $(\lambda\delta)(\lambda c)(\Sigma u : (\Sigma x : \texttt{Entity})\texttt{Man}(x))\texttt{Enter}(\pi_1(u))$

Term c represents the previous context and δ is its type. In (4), c and δ do not appear in the main clause of the representation; the context is not used in this example although it would be passed from the previous sentence.

In DTS, a semantic representation of a sentence given a particular context c and type δ is always of the type `type` [4]. Thus, given c and δ, (3) gives rise to the following judgment.

(18) $(\Sigma u : (\Sigma x : \texttt{Entity})\texttt{Man}(x))\texttt{Enter}(\pi_1(u)) : \texttt{type}$

DTS is based on the paradigm of the Curry-Howard correspondence, according to which propositions are identified with types; then the truth of a proposition is analyzed as the existence of a proof (i.e., proof-term) of the proposition. In other words, for any proposition (semantic representation) P of type `type`), we can say that P is true if and only if P is *inhabited*, that is, there exists a proof-term t such that $t : P$.

A proof-term for $(\Sigma x : A)B(x)$ is a *pair* (x, t) consisting of an object x of type A and a proof-term t of $B(x)$. Operator π_1 is a projection function that takes such a pair of objects and returns the first element of the pair; similarly, π_2 is a projection that returns the second element of a pair.

In the case of (11), a term u of type $(\Sigma x : \texttt{Entity})\texttt{Man}(x)$ is a pair consisting of a term x, which is of type `Entity`, and a proof-term of type $\texttt{Man}(x)$, which depends on the term x. Then, $\pi_1(u)$ represents its first projection, term x of type `Entity`. A proof-term for (11) consists of an entity x, a proof that x is a man, and a proof that x entered. Thus, sentence (3) is true if and only if there exists a tuple consisting of these objects.

Sentences with universal quantifiers like *every* are represented using Π type; for instance, (12) is represented as (13).

(19) Every man entered.

(20) $(\lambda\delta)(\lambda c)(\Pi u : (\Sigma x : \texttt{Entity})\texttt{Man}(x))\texttt{Enter}(\pi_1(u))$

A proof-term for $(\Pi x : A)B$ is a *function* such that for any object x of type A, it returns a proof-term t of $B(x)$. Thus, given a context c and its type δ, (12) is true if and only if there is a function such that for any pair u of an entity x and a proof-term of $Man(x)$, it returns a proof-term of $Enter(\pi_1(u))$. In other words, (12) means that if there exists an entity x and a proof that x is a man, then there exists a proof that x entered. Henceforth, `Entity` is abbreviated as `E` for simplicity. Also, we usually do not mention type δ to simplify notation.

Next, we explain how two sentences are connected in DTS. Consider the following example:

(21) $[\text{A man}]_i$ entered.

(22) He_i whistled.

The sentence (15) is represented as follows:

(23) $(\lambda c)\texttt{Whistle}(sel_{\texttt{E}}(c))$

The representation of the pronoun *he* is associated with a selection function sel. The selection function $sel_{\texttt{T}}(C)$ is a projection function or a composition thereof, which selects an appropriate antecedent of type `T` from the context C.

Dynamic conjunction is defined as follows:

(24) $P; Q \equiv (\lambda c)(\Sigma u : Pc)Q(c, u)$

In the first conjunct P, the information from the previous context is passed as an argument c. In the second conjunct Q, not only c but also the information from P, i.e., a proof-term u of type Pc, is passed as a pair (c, u). Thus, the semantic representation of the two sentences (14) and (15) is reduced to the following:

(25) $(\lambda c)(\Sigma v : (\Sigma u : (\Sigma x : \texttt{E})\texttt{Man}(x))\texttt{Enter}(\pi_1(u)))\texttt{Whistle}(sel_{\texttt{E}}((c, v)))$

Here, the proper choice of the selection function for the intended reading is such that $sel_{\texttt{E}}((c, v)) = \pi_1\pi_1\pi_2(c, v)$. This selects an entity x such that x is a man and x entered. In the framework of DTS, an element is accessible if and only if it can be taken from the context by the projections and the double-negation elimination rule. Thus, DTS is designed to handle dynamic binding as an inference in type theory. For more details, see [4].

4 Representation Given by Tanaka et al. (2013)

According to Tanaka et al. [20], noun phrases in restrictor position are represented in a uniform manner as follows:

Noun phrase in restrictor position	Representation in DTS
Farmers	$(\Sigma x : \mathtt{E})\mathtt{Farmer}(x)$
Farmers who own a donkey	$(\Sigma x : \mathtt{E})((\Sigma u : \mathtt{Farmer}(x))$ $(\Sigma y : \mathtt{E})(\Sigma w : \mathtt{Donkey}(y))\mathtt{Own}(x, y)))$
Farmers who own a donkey who are rich	$(\Sigma x : \mathtt{E})(\Sigma v : ((\Sigma u : \mathtt{Farmer}(x))$ $(\Sigma y : \mathtt{E})(\Sigma w : \mathtt{Donkey}(y))\mathtt{Own}(x, y))))\mathtt{Rich}(x)$

As we can see, the first projection of a given term for a restrictor always yields an object of type E, regardless of what the noun phrase in restrictor position is. Thanks to this, we can use type E as a fixed parameter for injection, surjection, and finiteness, in contrast to Sundholm's analysis we discussed in section 2.

Now *Most* A B for weak reading is defined as follows:

$$\frac{A : (\Pi x : \mathtt{E})(\Pi c : \delta)\mathtt{type} \quad B : (\Pi x : \mathtt{E})(\Pi c : \delta)\mathtt{type}}{\begin{array}{l} Most_{weak}(A, B) = (\lambda c)(\Sigma k : N)(k \geq [\pi_1(sel_{\mathtt{E}\text{-finite}((\Sigma x:\mathtt{E})Axc)}(c))/2] + 1 \\ \wedge(\Sigma f : M(k) \to (\Sigma x : \mathtt{E})Axc)(\mathtt{E}\text{-injection}(f) \wedge (\Pi y : M(k))(B(\pi_1(fy))(c, fy)))) \\ : (\Pi c : \delta)\mathtt{type} \end{array}}$$

Similarly to the case of dynamic conjunction we mentioned above, a pair (c, fy), where c is a previous context and fy encodes the information of A, is given as the second argument of B.

Tanaka et al. [20] consider other linguistic aspects of the determiner *most*, namely, donkey anaphora, the strong reading, and existential presupposition.

First, the internal anaphora in the donkey sentence (19) is analyzed as in (20):

(26) Most farmers who own [a donkey]$_i$ beat it$_i$.

(27) $(\lambda c)(\Sigma k : N)(k \geq [\pi_1(sel_{\mathtt{E}\text{-finite}(Res)}(c))/2] + 1$
$\wedge(\Sigma f : M(k) \to Res)(\mathtt{E}\text{-injection}(f) \wedge (\Pi y : M(k))(\mathtt{Beat}(\pi_1(fy), sel_{\mathtt{E}}((c, fy))))))$

Res stands for $(\Sigma x : \mathtt{E})(\mathtt{Farmer}(x) \wedge (\Sigma v : (\Sigma y : \mathtt{E})\mathtt{Donkey}(y))\mathtt{Own}(x, \pi_1 v)$, which corresponds to the restrictor *farmers who own a donkey*. In (13), the proper choice of $sel_{\mathtt{E}}$ is such that $sel_{\mathtt{E}}((c, fy)) = \pi_1\pi_2\pi_2\pi_2(c, fy)$, which leads to the intended interpretation.

They also discuss the *strong reading* [7] of donkey sentences. While the weak reading of the sentence (12) is that most farmers who own a donkey beat *at least one* donkey they own, the strong reading is that most farmers who own at least one donkey beat *every* donkey they own. Tanaka et al. [20] give a semantic representation of *most* that captures the strong reading as well:

$$\frac{A : (\Pi x : \mathtt{E})(\Pi c : \delta)\mathtt{type} \quad B : (\Pi x : \mathtt{E})(\Pi c : \delta)\mathtt{type}}{\begin{array}{l} Most_{strong}(A, B) = (\lambda c)(\Sigma k : N)(k \geq [\pi_1(sel_{\mathtt{E}\text{-finite}((\Sigma x:\mathtt{E})Axc)}(c))/2] + 1 \\ \wedge(\Sigma f : M(k) \to (\Sigma x : \mathtt{E})Axc)(\mathtt{E}\text{-injection}(f) \wedge (\Pi y : M(k))((B(\pi_1(fy))(c, fy) \\ \wedge(\Pi z : (\Sigma x : \mathtt{E})Axc)(eq(\mathtt{E}, \pi_1 z, \pi_1(f(y))) \to B(\pi_1 z)(c, z)))))) \\ : (\Pi c : \delta)\mathtt{type} \end{array}}$$

Owing to the addition of last conjunct, the first projection of any pair in the range of f is a farmer who beats every donkey he owns.

Finally, they take into account existential presupposition for *most*. *Most A B* presupposes that the set A is non-empty, as is usually assumed for strong determiners in general [10]. This presupposition can be predicted by modifying the selection function $sel_{\mathtt{E}\text{-finite}((\Sigma x:\mathtt{E})Axc)}$ in the definition of weak and strong *most* in such a way that *sel* returns a value only if the set of entities satisfying the restrictor is non-empty.

In sum, Tanaka et al. [20] provide an alternative definition for *Most A B* that solves the problem of uniformity remaining in the analysis by Sundholm [19]. Moreover, they consider other linguistic aspects of *most* and represent those features in their framework.

5 Proposal

Although Tanaka et al. [20] provide a definition for the uniformity of *most*, their definition leads to a wrong prediction in the following example:

(28) [Most farmers]$_i$ own a donkey. They$_i$ are rich.

In (10), *they* refers to the farmers who own a donkey [8]. Note that, this sentence is intuitively false in the following situation: there are 100 farmers, 80 of them own a donkey, and 60 farmers are rich. This means that *they* should be interpreted as *all farmers who own a donkey*. This appropriate interpretation, however, is not always guaranteed in the definition of Tanaka et al. [20]. According to their definition, k is greater than half of the cardinality of A and that there exist at least k elements satisfying both A and B. Therefore, k is not necessarily equal to the number of elements satisfying both A and B. If their definition is adopted, we cannot know the actual number of elements that satisfy both A and B, so the above example can be interpreted in the wrong way.

We provide the following revised representation for *most*.

$$\frac{A : (\Pi x : \mathtt{E})(\Pi c : \delta)\mathtt{type} \quad B : (\Pi x : \mathtt{E})(\Pi c : \delta)\mathtt{type}}{\begin{array}{l} Most_{weak}(A,B) = (\lambda c)(\Sigma k : N)(k \geq [\pi_1(sel_{\mathtt{E}\text{-finite}((\Sigma x:\mathtt{E})Axc)}(c))/2] + 1 \\ \wedge(\Sigma f : M(k) \rightarrow (\Sigma x : \mathtt{E})Axc)(\mathtt{E}\text{-injection}(f) \\ \wedge(\Pi z : (\Sigma x : \mathtt{E})Axc)(B(\pi_1 z)(c,z) \leftrightarrow (\Sigma y : M(k))eq((\Sigma x : \mathtt{E})Axc, z, fy))))) \\ : (\Pi c : \delta)\mathtt{type} \end{array}}$$

The last conjunct is different from the definition by Tanaka et al. [20]; in our new definition biconditional is used instead of implication in the last conjunct. It can be easily proved that the proposition defined here entails the one defined in [20]. The last conjunct in the revised definition means that there exists a one-to-one correspondence between the elements in $M(k)$ and the elements that satisfy both A and B. The existence of such a mapping f ensures that k is equal to the number of elements that satisfy both A and B, and now we know the actual number of *all farmers who own a donkey*.

We propose the following definition for the strong reading:

$$\frac{A : (\Pi x : \mathtt{E})(\Pi c : \delta)\mathtt{type} \quad B : (\Pi x : \mathtt{E})(\Pi c : \delta)\mathtt{type}}{\begin{array}{l} Most_{strong}(A, B) = (\lambda c)(\Sigma k : N)(k \geq [\pi_1(sel_{\text{E-finite}((\Sigma x:\mathtt{E})Axc)}(c))/2] + 1 \\ \wedge(\Sigma f : M(k) \to (\Sigma x : \mathtt{E})Axc)(\text{E-injection}(f) \\ \wedge(\Pi z : (\Sigma x : \mathtt{E})Axc)(B(\pi_1 z)(c, z) \leftrightarrow (\Sigma y : M(k))eq(\mathtt{E}, \pi_1 z, \pi_1(fy)))))) \\ : (\Pi c : \delta)\mathtt{type} \end{array}}$$

Instead of $eq((\Sigma x : \mathtt{E})Axc, z, fy)$ in the representation for the weak reading, $eq(\mathtt{E}, \pi_1 z, \pi_1(fy))$ is used in the representation for strong reading. This means that we disregard the proof-objects in elements of $(\Sigma x : \mathtt{E})Axc$ and only look at the first components of the pairs. Thus, for $Most_{strong}(A, B)$ to be true, all objects whose first projection is equal to $\pi_1(fy)$ must satisfy B. This enables counting the number of farmers that have a beating relation with respect to every donkey they own. This captures the intended meaning of the strong reading. Note that the description becomes simpler than that proposed by Tanaka et al. [20].

Our proposal can be naturally extended to numerical quantifiers such as *three, exactly three, at least three, more than three* and *fewer than three*. The respective representations for *Three A B*, *Exactly three A B* and *At least three A B* are as follows:

$$\frac{A : (\Pi x : \mathtt{E})(\Pi c : \delta)\mathtt{type} \quad B : (\Pi x : \mathtt{E})(\Pi c : \delta)\mathtt{type}}{\begin{array}{l} Three(A, B) = (\lambda c)(\Sigma k : N)(k = 3 \wedge (\Sigma f : M(k) \to (\Sigma x : \mathtt{E})Axc)(\text{E-injection}(f) \\ \wedge(\Pi z : (\Sigma x : \mathtt{E})Axc)((\Sigma y : M(k))eq((\Sigma x : \mathtt{E})Axc, z, fy) \to B(\pi_1 z)(c, z))))) : (\Pi c : \delta)\mathtt{type} \end{array}}$$

$$\frac{A : (\Pi x : \mathtt{E})(\Pi c : \delta)\mathtt{type} \quad B : (\Pi x : \mathtt{E})(\Pi c : \delta)\mathtt{type}}{\begin{array}{l} Exactly\ three(A, B) = (\lambda c)(\Sigma k : N)(k = 3 \wedge (\Sigma f : M(k) \to (\Sigma x : \mathtt{E})Axc)(\text{E-injection}(f) \\ \wedge(\Pi z : (\Sigma x : \mathtt{E})Axc)(B(\pi_1 z)(c, z) \leftrightarrow (\Sigma y : M(k))eq((\Sigma x : \mathtt{E})Axc, z, fy)))) : (\Pi c : \delta)\mathtt{type} \end{array}}$$

$$\frac{A : (\Pi x : \mathtt{E})(\Pi c : \delta)\mathtt{type} \quad B : (\Pi x : \mathtt{E})(\Pi c : \delta)\mathtt{type}}{\begin{array}{l} At\ least\ three(A, B) = (\lambda c)(\Sigma k : N)(k \geq 3 \wedge (\Sigma f : M(k) \to (\Sigma x : \mathtt{E})Axc)(\text{E-injection}(f) \\ \wedge(\Pi z : (\Sigma x : \mathtt{E})Axc)(B(\pi_1 z)(c, z) \leftrightarrow (\Sigma y : M(k))eq((\Sigma x : \mathtt{E})Axc, z, fy)))) : (\Pi c : \delta)\mathtt{type} \end{array}}$$

The difference between *three* and *exactly three* is that the former is defined in terms of implication, while the latter is defined in terms of biconditional. This ensures that *Three A B* is true when there are more than three elements which satisfy both A and B, while *exactly three* is not true in the same situation.

The sentences *Three A B* and *At least three A B* have the same truth condition, but they are defined in a different way. This is because they behave differently when the subjects are referred to from the subsequent sentences. As Kadmon [11] observed, *three A* differs from *at least three A* in anaphora possibilities. For instance, in (29), *they* refers to a collection of exactly ten kids. By contrast, in (35), *they* refers to the maximal collection of kids who walked into the room. Thus, if there are twelve kids who walked into the room, *they* in (35) can refer to all the twelve kids, while *they* in (29) cannot.

(29) [Ten kids]$_i$ walked into the room. They$_i$ were making an awful lot of noise.

(30) [At least ten kids]$_i$ walked into the room. They$_i$ were making an awful lot of noise.

In our definition, $Three(A, B)$ is defined with the cardinality condition $k = 3$, whereas *At least three*(A, B) requires the condition $k \geq 3$. Furthermore, in the definition of *At least three*(A, B), biconditional rather than implication is used in the last conjunct. This captures the difference in anaphora possibilities between (29) and (35).

More than three A B and *Fewer than three A B* can be defined in the same way as *Exactly three A B* and *At least three A B*. All we have to do is use respectively $k > 3$ or $k < 3$ as a condition of k. In this way, our definition captures the difference between various cardinal quantifiers in a perspicuous way. By contrast, in the theory of Tanaka et al. [20], more substantial redefinitions are required to accommodate downward monotonic and non-monotonic determiners, thus resulting in different forms of definition for different types of determiners.

6 Conservativity and Right Upward Monotonicity

In this section, we first show that our definition for *most* satisfies conservativity. Conservativity is a property of determiners which is model-theoretically formulated as: for all M and $X, Y \subseteq M$, $Q_M(X, Y) \Leftrightarrow Q_M(X, X \cap Y)$ (see e.g., [16]). For our purpose, conservativity property must be defined from a proof-theoretic perspective. Furthermore, since our formalization of GQs in DTS contains a mechanism for context passing, we need to formulate conservativity property in our dynamic settings. We define the conservativity property using dynamic conjunction $P; Q$ introduced in section 3.

Definition 1 (Conservativity) For any $A, B : (\Pi x : \mathtt{E})(\Pi c : \delta)\mathtt{type}$, a determiner Q is conservative with respect to a context c if the following holds:

$$Q(A, B)c \text{ is inhabited} \Longleftrightarrow Q(A, (\lambda x)(Ax; Bx))c \text{ is inhabited.}$$

Note that $Q(A, B)$ is of the type $(\Pi c : \delta)\mathtt{type}$, that is, a function from a context to a proposition, so we assume that both $Q(A, B)$ and $Q(A, (\lambda x)(Ax; Bx))$ are given the same context c. For the current purpose of this paper, we focus on the case of the determiner *most* in its weak reading. It is straightforward to generalize the proof to other determiners we considered so far. Assuming that a context c is fixed, $Most_{weak}(A, B)c$ and $Most_{weak}(A, (\lambda x)(Ax; Bx))c$ are defined as follows.

$Most_{weak}(A, B)c$	$Most_{weak}(A, (\lambda x)(Ax; Bx))c$
$(\Sigma k : N)$	$(\Sigma k : N)$
$(k \geq [\pi_1(sel_{\mathtt{E}\text{-finite}((\Sigma x:\mathtt{E})Axc)}(c))/2] + 1$	$(k \geq [\pi_1(sel_{\mathtt{E}\text{-finite}((\Sigma x:\mathtt{E})Axc)}(c))/2] + 1$
$\wedge(\Sigma f : M(k) \to (\Sigma x : \mathtt{E})Axc)$	$\wedge(\Sigma f : M(k) \to (\Sigma x : \mathtt{E})Axc)$
$(\mathtt{E}\text{-injection}(f) \wedge (\Pi z : (\Sigma x : \mathtt{E})Axc)$	$(\mathtt{E}\text{-injection}(f) \wedge (\Pi z : (\Sigma x : \mathtt{E})Axc)$
$(B(\pi_1 z)(c, z)$	$((\Sigma u : A(\pi_1 z)(c, z))B(\pi_1 z)((c, z), u)$
$\leftrightarrow (\Sigma y : M(k))eq((\Sigma x : \mathtt{E})Axc, z, fy))))$	$\leftrightarrow (\Sigma y : M(k))eq((\Sigma x : \mathtt{E})Axc, z, fy))))$

In the following proof, we use introduction and elimination rules for Σ-type and Π-type. For these rules, see e.g., [15] and [17].

Theorem 1 $Most_{weak}$ is conservative with respect to a given context c, that is, $Most_{weak}(A, B)c$ is inhabited $\Longleftrightarrow$ $Most_{weak}(A, (\lambda x)(Ax; Bx))c$ is inhabited.

Proof. We show the derivation only for the left-to-right direction. The derivation for the right-to-left direction can be given in a similar way. Let m be a proof-term of $Most_{weak}(A, B)c$. For the left-to-right direction, it suffices to show that the following is inhabited.

$$\begin{array}{l}(\Pi z : (\Sigma x : \mathtt{E})Axc)((\Sigma u : A(\pi_1 z)(c, z))B(\pi_1 z)((c, z), u) \\ \qquad\qquad \leftrightarrow (\Sigma y : M(\pi_1 m))eq((\Sigma x : \mathtt{E})Axc, z, (\pi_1\pi_2\pi_2 m)y))))\end{array} \quad (1)$$

So assume that $z : (\Sigma x : \mathtt{E})Axc$. To show one direction in (1), assume the following:

$$d : (\Sigma u : A(\pi_1 z)(c, z))B(\pi_1 z)((c, z), u). \quad (2)$$

By applying Σ-elimination several times to $m : Most_{weak}(A, B)$, we obtain:

$$\begin{array}{l}\pi_2\pi_2\pi_2\pi_2 m : (\Pi z : (\Sigma x : \mathtt{E})Axc)(B(\pi_1 z)(c, z) \\ \qquad\qquad \leftrightarrow (\Sigma y : M(\pi_1 m))eq((\Sigma x : \mathtt{E})Axc, z, (\pi_1\pi_2\pi_2 m)y).\end{array} \quad (3)$$

By applying Π-elimination with $z : (\Sigma x : \mathtt{E})Axc$ and (3) and then applying Σ-elimination, we have:

$$\pi_1((\pi_2\pi_2\pi_2\pi_2 m)z) : B(\pi_1 z)(c, z) \rightarrow (\Sigma y : M(\pi_1 m))eq((\Sigma x : \mathtt{E})Axc, z, (\pi_1\pi_2\pi_2 m)y). \quad (4)$$

By Σ-elimination with (2), we get

$$\pi_2 d : B(\pi_1 z)((c, z), \pi_1 d),$$

from which the following can be derived by an admissible rule whose proof is omitted here due to space limitation. [4]

$$\pi_2 d : B(\pi_1 z)(c, z) \quad (5)$$

Then by Π-elimination with (4) and (5), we have:

$$(\pi_1((\pi_2\pi_2\pi_2\pi_2 m)z))(\pi_2 d) : (\Sigma y : M(\pi_1 m))eq((\Sigma x : \mathtt{E})Axc, z, (\pi_1\pi_2\pi_2 m)y). \quad (6)$$

Thus, by Π-introduction with (2) and (6), we have:

$$\begin{array}{l}(\lambda d)(\pi_1((\pi_2\pi_2\pi_2\pi_2 m)z))(\pi_2 d) : \\ \qquad (\Sigma u : A(\pi_1 z)(c, z))B(\pi_1 z)((c, z), u) \rightarrow (\Sigma y : M(\pi_1 m))eq((\Sigma x : \mathtt{E})Axc, z, (\pi_1\pi_2\pi_2 m)y).\end{array} \quad (7)$$

To show the other direction in (1), assume:

$$e : (\Sigma y : M(\pi_1 m))eq((\Sigma x : \mathtt{E})Axc, z, (\pi_1\pi_2\pi_2 m)y). \quad (8)$$

[4] More specifically, we can prove that if a dynamic proposition M inhabits a term m under a context (c, z) and M is given a context c, where z is not used in M, then the proposition M' that is obtained by substituting each selection function of the form $\pi_1 f$ with f inhabits a term m' as well. A similar rule is used to obtain (12): a rule that if a dynamic proposition M inhabits a term m under a certain context c and M is given a pair of context (c, z), where z is not used in M, then the proposition M' that is obtained by substituting each selection function f with $f \circ \pi_1$ inhabits a term m' as well.

In a similar way to derive (4) above, we have:

$$\pi_2((\pi_2\pi_2\pi_2\pi_2m)z) : (\Sigma y : M(\pi_1m))eq((\Sigma x : \mathtt{E})Axc, z, (\pi_1\pi_2\pi_2m)y) \to B(\pi_1z)(c, z). \quad (9)$$

Then by Π-elimination with (14) and (15), we have:

$$(\pi_2((\pi_2\pi_2\pi_2\pi_2m)z))e : B(\pi_1z)(c, z). \quad (10)$$

By applying Σ-elimination with $z : (\Sigma x : \mathtt{E})Axc$, we have:

$$\pi_2z : A(\pi_1z)c. \quad (11)$$

Then, by applying the admissible rule mentioned above to (10) and (17), we derive:

$$\pi_2z : A(\pi_1z)(c, z) \text{ and } (\pi_2((\pi_2\pi_2\pi_2\pi_2m)z))e : B(\pi_1z)((c, z), \pi_2z). \quad (12)$$

So, by Σ-introduction with (12), we get:

$$(\pi_2z, \pi_2\pi_1((\pi_2((\pi_2\pi_2\pi_2\pi_2m)z))e)) : (\Sigma u : A(\pi_1z)(c, z))B(\pi_1z)((c, z), u). \quad (13)$$

Then by Π-introduction with (14) and (19), we get:

$$\begin{array}{l}(\lambda e)(\pi_2z, \pi_2\pi_1((\pi_2((\pi_2\pi_2\pi_2\pi_2m)z))e) : \\ \quad (\Sigma y : M(\pi_1m))eq((\Sigma x : \mathtt{E})Axc, z, (\pi_1\pi_2\pi_2m)y) \to (\Sigma u : A(\pi_1z)(c, z))B(\pi_1z)((c, z), u).\end{array} \quad (14)$$

Therefore, by Σ-introduction with (7) and (14) and by Π-introduction with $z : (\Sigma x : \mathtt{E})Axc$, we can obtain a proof-term for (1), as required. □

We can also formulate monotonicity properties in our setting. The definition of right upward (downward) monotonicity is as follows.

Definition 2 (Monotonicity) For any $A, B, B' : (\Pi x : \mathtt{E})(\Pi c : \delta)\mathtt{type}$, a determiner Q is right upward (resp. downward) monotonic given a context c if both $(\Pi x : \mathtt{E})(Bxc \to B'xc)$ (resp. $(\Pi x : \mathtt{E})(B'xc \to Bxc)$) and $Q(A, B)c$ are inhabited $\Longrightarrow Q(A, B')c$ is inhabited.

Left upward and downward monotonicity can be defined in an obvious way. As for determiners we discussed so far, it can be proved that (i) right upward monotonic holds for *most, three, at least three, more than three*; (ii) left upward monotonicity holds for *three, at least three, more than three*; (iii) left and right downward monotonicity holds for *fewer than three*. We note that a proof of monotonicity for *exactly three* is properly blocked. For the space limitation, we omit the proofs here.

7 Conclusion

In this paper, we presented the formalization of GQs in the framework of DTS. We provided an appropriate and simplified semantic representation for the determiner *most* and extend the approach to numerical quantifiers. We also proved that the GQs we defined here satisfy conservativity and monotonicity property in our dynamic setting.

It is interesting to see whether the formalization proposed here can be naturally extended to three-place determiners (type $\langle\langle 1,1\rangle,1\rangle$ determiners) such as *at least as many A as B are C* and *more A than B are C*. The development and analysis for such more complicated GQs remain for future work.

References

1. Aczel, P.: The Type Theoretic Interpretation of Constructive Set Theory: Choice Principles. In: Troelstra, A.S., van Dalen, D. (eds.) The L.E.J Brouwer Centenary Symbosium. Amsterdam (1982)
2. Barendregt, H.P.: Lambda Calculi with Types. In: Handbook of logic in computer science, vol. II, pp. 117–309. Oxford University Press (1992)
3. Barwise, J., Cooper, R.: Barwise and Cooper 1981.pdf. In: Linguistics and philosophy, Vol. 4, No. 2, pp. 159–219. Springer (1981)
4. Bekki, D.: Dependent Type Semantics: An Introduction. In: Zoe, C., Galeazzi, P., Gierasimczuk, N., Marcoci, A., Smets, S. (eds.) Logic and Interactive RAtionality Yearbook 2012, Vol.1, pp. 277–300. University of Amsterdam (2014)
5. Bekki, D.: Representing Anaphora with Dependent Types. In: Asher, N., Soloviev, S. (eds.) Logical Aspects of Computational Linguistics, pp. 14–29. Springer Berling Heidelberg (2014)
6. van Benthem, J.: A brief history of natural logic. In: Chakraborty, M., Löwe, B., Nath Mitra, M. (eds.) Logic, Navya-Nyaya & Applications, Homage to Bimal Krishna Matilal. College Publications, London (2008)
7. Chierchia, G.: Anaphora and Dynamic Binding. Linguistics and philosophy 15, 111–183 (1992)
8. Evans, G.: Pronouns. Linguistic Inquiry 11(2), 337–362 (1980)
9. Francez, N., Ben-Avi, G.: Proof-Theoretic Reconstruction of Generalized Quantifiers. Journal of Semantics pp. 1–59 (2014)
10. Heim, I., Kratzer, A.: Semantics in Generative Grammar. Blackwell (1998)
11. Kadmon, N.: On Unique and Non-unique Reference and Asymmetric Quantification (1992)
12. Kamp, H., Reyle, U.: From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory, vol. 1. Kluwer Academic Publisher (1993)
13. Luo, Z.: Formal semantics in modern type theories with coercive subtyping. Linguistics and Philosophy 35(6), 491–513 (2013)
14. Martin-Löf, P.: An Intuitionistic Theory of Types: Predicative Part. Studies in Logic and the Foundations of Mathematics 80, 73–118 (1975)
15. Martin-Löf, P.: Intuitionistic Type Theory. Bibliopolis Naples (1984)
16. Peters, S., Westerståhl, D.: Quantifiers in Language and Logic. Clarendon Press (2006)
17. Ranta, A.: Type-theoretical Grammar. Oxford University Press (1995)
18. Sundholm, G.: Proof Theory and Meaning. In: Handbook of Philosophical Logic, pp. 471–506. Springer Netherlands (1986)
19. Sundholm, G.: Constructive Generalized Quantifiers. Synthese 79(1), 1–12 (1989)
20. Tanaka, R., Nakano, Y., Bekki, D.: Constructive Generalized Quantifiers Revisited. In: Nakano, Y., Satoh, K., Bekki, D. (eds.) New Frontiers in Artificial Intelligence (JSAI-isAI 2013 Workshops, LENLS, JURISIN, MiMI, AAA, and DDS, Kanagawa, Japan, October 27-28, 2013, Revised Selected Papers). Springer Berling Heidelberg (to appear)

Semantics of *to Count*[*]

Dina Voloshina

Goethe University Frankfurt
dina.voloshina@outlook.de

Abstract In "Intensional Verbs and Quantifiers" (1997) Friederike Moltmann suggests new linguistic tests to reveal intensional verbs. According to the new tests, the verb **to count** comes out as intensional. In this paper I will question the new tests. I will show that they do not work as they are supposed to work. Further, we will see that both traditional and new criteria reveal not only classical intensional verbs, but also measure verbs as **to weigh** or **to measure**. We will find similarities in behavior between measure verbs and **to count**. On the basis of these similarities, I suggest to analyze the non-extensional **to count** as a measure verb with a measure phrase as object.

1 Introduction

In this paper I am concerned with the question of how to analyze the ambiguity of **to count**. I will argue against Friederike Moltmann's analysis in "Intensional Verbs and Quantifiers" (1997). She suggests to analyze the 2 (or even 3) readings of sentences containing **to count** as the extensional/intensional distinction:

(1) **John counted 28 ships.**

1 the extensional reading: there was a group of 28 ships, and John counted them (the result may be unknown to the speaker);
2 the number-intensional reading: there was a group of ships, which John counted coming up with the result 28;
3 the (possible) sortal-intensional reading: there was a group of objects, which John falsely identified as ships and counted coming up with the result 28.[1]

Moltmann argues for intensionality of **to count** by means of the traditional and new criteria for intensionality, which she suggests. However, we will see that criteria for intensionality reveal both intensional verbs and measure verbs.

I argue for an alternative analysis of **to count**: The source of the ambiguity is not the opacity, but a lexical ambiguity, a polysemy between an extensional

[*] I would like to thank Ede Zimmermann – the supervisor of my Master's Thesis, on which this paper is based – for all his help and guidance.

[1] Based on the original analysis of Moltmann, I will call these readings number-intensional and sortal-intensional for reasons of reference only. I do not necessarily assume their intensionality.

to count$_{\mathrm{t}}$[2] and a measure verb **to count**$_{\mathrm{m}}$ (similar to measure verbs like **to weigh** or **to measure**). Both so-called intensional readings are different usages (or – depending on the analysis – readings) of **to count**$_{\mathrm{m}}$.

2 The Traditional Criteria for Intensionality Applied to *to Count*

Below we see the traditional criteria for intensionality (cf. Zimmermann (2001)):[3]

- Existential Impact
  From *x Rs an N* infer: *There is at least one N.*
- Extensionality
  From *x Rs an N*, *Every N is an M* und *Every M is an N* infer:
  x Rs an M.

Extensional verbs like **to kiss** or **to meet** satisfy both criteria. Intensional verbs like **to look for** violate at least one criterion. What about **to count**?

First let's consider the number-intensional reading of the sentence in (1) in a situation in which **ship** and **yacht** are co-extensional. Extensionality criterion is fulfilled, but not that of Existential Impact. (2) follows from (1), but not (3):

(2) $\Rightarrow$**John counted 28 yachts.**
(3) $\not\Rightarrow$ **There were 28 ships.**

Now let's consider the sortal-intensional reading in a situation in which there are 28 icebergs on the sea and John identified them as ships. **Ship** and **yacht** are co-extensional. Neither Extensionality nor Existential Impact is satisfied:

(4) $\not\Rightarrow$**John counted 28 yachts.**
(5) $\not\Rightarrow$ **There were 28 ships.**

At first glance, **to count** looks like an intensional verb. However, we will see that according to the criteria for intensionality not only classical intensional verbs, but also measure verbs come out as intensional.

3 Similarity with Measure Phrases

Moltmann (1997) admits that there might be a difference between the intensional object of **to count** and ordinary quantifiers: the object of **to count** seems to function like a measure phrase in measure constructions, which also fail to satisfy the Existential Impact criterion. We can see it on the example below cited from Moltmann (1997, p.45):

(6) **The box weighs at least two kilos.**

[2] with **t** standing for *transparent*
[3] I do not consider the criterion of Specificity, which is intended to apply to singular indefinites.

(7) ⇏**There are at least two kilos x such that the box weighs x.**

I argue that the object of the so-called intensional **to count** is a measure phrase. I do not assume an intensional account for the object of **to count** in the number-intensional reading or for other measure phrases.

Let's look at examples which provide evidence for a different semantic status of measure phrases. The german so-called container expressions (a kind of measure phrases) like **zwei Becher Milch** (two cup milk, *2 cups of milk*) or **drei Glas/Gläser Wein** (three glass-SG/glass-PL wine, *three glasses of wine*) show an ambiguity between a concrete-objects reading in (8) and a measure-phrase reading in (9):

Margot hat viele Gläser gespült.
(8) Margot PAST many glasses wash-PTCP
Margot washed many glasses.
Margot hat viele Gläser getrunken.
(9) Margot PAST many glasses drink-PTCP
Margot drank many glasses.

I claim that the ambiguity of **to count** is of the same nature as the ambiguity of container expressions, i.e. the ambiguity between the concrete-objects reading and the measure-phrase reading. The source of the ambiguity is not the opacity.

Just as the objects of the alleged intensional **to count**, container expressions don't satisfy the Existential Impact criterion:

Helmut hat gestern ein Päckchen Zigaretten geraucht
(10) Helmut PAST yesterday a pack cigarettes rauchen-PTCP
Yesterday Helmut smoked a pack of cigarettes.
(11) ⇏**There was a pack of cigarettes, which Helmut smoked.**

(11) does not follow from (10): Helmut could smoke cigarettes from different packs. We can assert (10), as long as the amount of smoked cigarettes measures up to the amount of cigarettes in a pack (cf. Zifonun et al. 1998, p.1985).

The analysis of the object of the number-intensional **to count** as a measure phrase would explain yet another peculiarity, i.e. the fact that the number-intensional **to count** – contrary to the transparent reading – allows only numerals as objects. (12) has only a transparent reading:

(12) **John counted ships.**

This conforms to the observation that numerals are obligatory with container expressions in the measure phrase reading and other measure phrases:

(13) **#The book weighs this pound.**

4 The New (Linguistic) Criteria for Intensionality

The problem with the traditional criteria for intensionality is that the results are not always clear and uncontroversial. Moltmann (1997) suggested new criteria, which should be more decisive. Let's take a closer look at them.

4.1 Lack of Anaphora Support

In general, anaphora support is impossible with the intensional reading of an intensional verb:

(14) (a) **John is looking for a horse. Mary is looking for it too.**
(b) **John is looking for a horse. It must be white and have a golden mane.**

In (14a) **to look** has only extensional reading. Anaphora support with the intensional reading of an intensional verb is only possible in a modal context as in (14b) (cf. Moltmann 1997, p.6).

4.2 Use of Impersonal Proforms

The use of the proform for objects of extensional verbs – whether personal or impersonal – is dependent on the corresponding NP. According to Moltmann, intensional verbs may only use impersonal proforms for both animate and inanimate objects (in the examples below the verb **to look for** is used intensionally; the ungrammaticality signs # and **??** refer only to the intensional reading of **to look for**):

(15) (a) **John is looking for something, namely a secretary.**
(b) **#John is looking for someone, namely a secretary.**
(c) **#John met something, namely a secretary.**
(d) **John met someone, namely a secretary.**
(16) (a) **What is John looking for?- A secretary.**
(b) **#Whom is John looking for?- A secretary.**
(c) **#What did John meet?- A secretary.**
(d) **Whom did John meet?- A secretary.**
(17) (a) **John is looking for two things, a secretary and an assistant.**
(b) **??John is looking for two people, a secretary and an assistant.**
(c) **#John met two things, a secretary and an assistant.**
(d) **John met two people, a secretary and an assistant.**

The evaluations of examples are cited from Moltmann's paper. However, not all native speakers agree with them, taking issue with the evaluations of all the (b)-examples. Contrary to the ungrammaticality signs # and **??**, these examples seem to be even better than the (a)-examples. We therefore have to weaken the test. The new version should be: extensional verbs allow only personal while intensional verbs allow both personal and impersonal proforms for their animate objects.[4]

[4] Independent of the extensionality/intensionality of the verbs, the unspecific reading in the (b)-examples should be possible according to (*):

* *Upward monotonicity:*
x is looking for a P.
$\Rightarrow$ x is looking for a Q,

where Q is a more general term than P, i.e. that the extension of P is a subset of the extension of Q (cf. Zimmermann 2006, p.721).

4.3 Identity Conditions

According to Moltmann, object-NPs of intensional verbs can – in contrast to extensional verbs – have the same semantic value, even if the intentional objects are different. It is possible to say (18) and (19), even if John and Mary are going to hire different assistants:

(18) **John is looking for the same thing as Mary, namely a new assistant.**
(19) **John is looking for what Mary is looking for, namely a new assistant.**

That shouldn't work with extensional verbs:

(20) **#John met the same thing as Mary, namely a new assistant.**
(21) **#John met what Mary met, namely a new assistant.**

However, the example in (20) chosen by Moltmann is not suitable. (20) must be bad for some other reason: as we have seen in section 4.2, the extensional verb **to meet** does not allow impersonal proforms. To demonstrate the test we need an inanimate object. In contrast with the predictions of the test, (22) is fine:

(22) **John bought the same thing as Mary, namely a new car.**

The free-relative version of the test does not work in the next example either. (23) is a good English sentence, even if John and Mary ate different apples:

(23) **John ate what Mary ate.**

It seems that something different is being tested here.

5 *To Count* and the New Criteria for Intensionality

Using the new tests, Moltmann suggests two new groups of intensional verbs: epistemic verbs (**to distinguish**, **to recognize**, **to discriminate**, **to count**, in addition to the already known ones **to see**, **to feel**, **to hear**) and resultative verbs (**to appoint**, **to hire**, **to elect**, **to choose**, **to find**). We will look at how the number-intensional **to count** behaves in the new tests and compare it with the measure verbs. That will provide us with further evidence for considering the object of the number-intensional **to count** a measure phrase.

5.1 *To count* and Anaphora Support

Anaphora support is impossible with the intensional **to count**. (24) has no number-intensional reading:

(24) **John counted 10 ships and Bill counted them too (though there were actually 12 ships).** (cf. Moltmann 1997, p.44)

Interestingly, the same holds for measure verbs. (25) has no such a reading as "John measured 3 meters and Bill measured 3 meters.":

(25) **John measured 3 meters and Bill measured them, too.**

Apparently, the test reveals measure phrases as well.

5.2 *To count* and the Proforms

We have seen in section 4.2 that intensional verbs allow both personal and impersonal proforms for their animate objects. Surprisingly, non-extensional **to count** behaves differently with the different versions of this test. Unlike classical intensional verbs, both non-extensional readings of **to count** allow only the impersonal proform as an interrogative particle:

(26) **What/#Whom did John count? – 10 men and 15 women.**
(cf. Moltmann 1997, p.44)

In (26) the evaluations of grammaticality refer only to the non-extensional readings of the verb.

In an affirmative sentence, the non-extensional **to count** seems to allow both personal and impersonal proforms:

(27) **John counted something/some people, namely 10 men and 15 women.**

However, I believe that **some people** is a measure phrase in this example – just as **10 men and 15 women**, only more general – different in type from the plural proform **two people** in (17d). Being a more general measure phrase, it must be allowed in this sentence for reason of monotonicity. If we change the proform the way it cannot function as a measure phrase – for example, by leaving out the determiner – we force the extensional reading of the object:

(28) **John counted people, namely 10 men and 15 women.**
(29) **John is looking for people (and not animals), namely a secretary and an assistant.**

Whereas (29) still has an intensional reading, the intensional reading has vanished in (28).

We will now see that measure verbs behave just the same! Consider the following situation. A new car has been constructed. It should conform to the contemporary standards: it should be small, but efficient. The back seat of the car is being checked to find out whether three people fit. John, the manager, asks three of his colleagues to sit on the back seat. Unfortunately there is place only for two of them. John says to the engineers reproachfully:

(30) **I only measured two people.**

The proform test applied to (30) delivers the following results (with the evaluations of grammaticality referring only to the intransitive **to measure**):

(31) **John measured something, namely two people.**
(32) **John measured some people, namely two.**
(33) **#John measured people, namely two.**
(34) **What did John measure? – Two people.**
(35) **#Whom did John measure? – Two people.**

We see in the exampeles above that the personal proforms are as bad with the measure verbs as with the non-extensional **to count**.

The proform test shows that there is a difference between intensional verbs on the one hand and **to count** and measure verbs on the other hand: whereas the former always allow both personal and impersonal proforms, the latter allow personal proforms only in cases which are subject to monotonicity.

5.3 *To count* and Identity Conditions

According to the free-relative version of the test, **to count** should be intensional:

(36) **John counted what Mary counted, namely 10 men and 15 women.**

Let's try the **the-same-thing** version of the test. For some reason, **the-same-thing** sentences with the number- or sortal-intensional **to count** are worse than the examples with the intensional **to look for** as in (18). The intuition is not very clear, but (38) seems to be much better than (37):

(37) **?John counted the same thing as Mary, namely 10 men and 15 women.**
(38) **John counted the same as Mary, namely 10 men and 15 women.**

Given that there was no such contrast between the two versions of the test with the intensional **to look for**, the contrast in acceptability between (37) and (38) shows that there is something different about the object of non-extensional **to count**, which makes accepting (37) more difficult.

Again, measure verbs behave exactly the same:

(39) **John weighed what Mary weighed, namely 2 kilos.**
(40) **?John weighed the same thing as Mary, namely 2 kilos.**
(41) **John weighed the same as Mary, namely 2 kilos.**

Whereas (39) and (41) are totaly fine, (40) is more difficult to accept.[5]

Why is the proform **the same thing** much worse with **to count** and with measure verbs? Having made an additional assumption about the presupposition of **thing**, we will explain it within my analysis of **to count** as a measure verb. We have to assume that apparently **thing** presupposes an individual domain,

[5] I thank an anonymous reviewer for pointing out that the examples in (37) and (40) are not impossible, contrary to what I used to assume and what F.Moltmann assumes in Moltmann (2013, p.166). Real-life examples provided by the reviewer are:

* - **Pats had only 10 men on the field.**
- **I counted the same thing.**
(from https://twitter.com/MikeReiss/status/120916160689606656)

* **Weighed myself the other day. I weighed the same thing as last time.**
(from https://twitter.com/seauxbreezy/status/379330551376646144)

which excludes not only animate objects, but measure values as well. Being measure values, the objects of the number-intensional and sortal-intensional **to count** and of measure verbs are generally not allowed. Some kind of coercion must be involved here that allows the combination of **the same thing** and measure phrases for some speakers.[6]

This test shows one more time that **to count** behaves neither as a classical intensional verb nor as a classical extensional verb.

6 Analysis of *to Count* as a Measure Verb

The core of my analysis consists in the assumption that the objects of both of the so-called intensional readings of **to count** are *measuring term phrases* (the expression comes from Thomason (1979), who also defends an extensional analysis of the objects of measure verbs). Measuring term phrases differ in type from the ordinary quantifiers of type $(et)t$. Counting is measuring (among authors who also defend this view are Wiese (1995), Eschenbach (1995)).

The difference between the extensional verb **to count**$_{\mathrm{t}}$ and the measure verb **to count**$_{\mathrm{m}}$ consists in the different functions of their respective objects. The object of **to count**$_{\mathrm{t}}$ is the patient, i.e. that which is to be counted, whereas the object of **to count**$_{\mathrm{m}}$ is a measure value.

The following are the questions that remain to be solved: 1) is the sortal-intensional reading a genuine reading? 2) how are the number-intensional reading and the sortal-intensional reading related?

I would like to propose two kinds of answers to these questions.

6.1 Pragmatic Solution

It is unclear whether the sortal-intensional reading is a genuine reading at all. That's why we could assume that the number-intensional reading is the only reading of **to count**$_{\mathrm{m}}$ and that the alleged sortal-intensional reading is a pragmatic phenomenon. Given that there are no ships on the sea, we have to reinterpret **John counted 28 ships** and conclude that John must have mistaken the icebergs for ships. It's not a new idea to offer a pragmatic explanation for alleged intensional epistemic readings. Montague assumed for verbs of perception like **to see** that they only have an extensional reading. The alleged intensional epistemic reading is only a pragmatic phenomenon, a reinterpretation of the extensional reading. Due to the fact that no unicorns exist, the only semantically available reading of **John sees a unicorn** – namely that there exists a unicorn which John sees – comes out as false and the sentence thus has to be reinterpreted. The hearer concludes that John apparently seems to see a unicorn (Montague 1974, p.169f).

[6] That would be in accordance with one of my informants who feels the denotation of **2 kilos** in (40) shift from a measure value to an individual.

With the pragmatic explanation of the sortal-intensional reading there remain only the transparent reading and the number-intensional reading. [7]

I assume that **to count**$_{\mathrm{t}}$ and **to count**$_{\mathrm{m}}$ have an event argument $\boldsymbol{\varepsilon}$. $\boldsymbol{\varepsilon}$ denotes a *completed* act of counting. The act of counting must be completed because I assume that we can only assert that someone counted something if they have finished counting. The counting result may be unknown to the speaker. Further, I introduce a constant **C** which, applied to a situation, denotes a three-place relation (with the transparent **to count** in the metalanguage):

(42) $[\![\mathbf{C}]\!] = \lambda s.\lambda y^*.\lambda x.\lambda\varepsilon$. the counting person x counts the group of objects y^* in the situation s in the act of counting ε

(42) says that a counting event, an ivdividual and a group are related by $\mathbf{C}_i$ if this individual has counted the group as the result of the counting event. That the counting person has finished counting, is guaranteed by the fact that the act of counting is completed.

A sentence with **to count**$_{\mathrm{t}}$ translates as follows:

(43) $\quad$ |**John counted**$_{\mathrm{t}}$**10 ships**|
$\equiv$ $\mathbf{C}_{\boldsymbol{i}}(\boldsymbol{\varepsilon}$**, john′, 10 ships′**$)$,

(43) says that a completed act of counting took place, in which John was the counting one and the group of 10 ships was that which was counted.

Accordingly the contribution of **to count**$_{\mathrm{t}}$ is:

(44) $(\boldsymbol{\lambda y}.(\boldsymbol{\lambda x}.(\boldsymbol{\lambda\varepsilon}.\mathbf{C}_{\boldsymbol{i}}(\boldsymbol{\varepsilon}, \boldsymbol{x}, \boldsymbol{y}))))$,

where $\boldsymbol{\varepsilon}$ is an act-of-counting variable, $\boldsymbol{x}$ is an individual variable and $\boldsymbol{y}$ is a plural-individual variable.

Let's come to **to count**$_{\mathrm{m}}$. I believe that at least in the number-intensional reading it is not intensional. The alleged intensionality comes from the false assumption that the counting result is an assertion about the actual number of counted objects. I think that the counting result is a claim only about the result itself. One can count the same group of objects many times with different results. And one is still justified in claiming that the first time one counted n objects, second time one counted m objects and so on. **To count**$_{\mathrm{m}}$ describes only the counting process and the respective counting result.It does not make any assertion about the actual number of the objects. The objects to be counted – however many they may be – are an implicit argument: it is given by context, which group of objects is to be counted.

I introduce a function **f**. **f** is a result function, which is only defined for acts of counting and which assigns a counting result $\boldsymbol{n}$ to each act of counting $\boldsymbol{\varepsilon}$: $\mathbf{f}(\boldsymbol{\varepsilon}) = \boldsymbol{n}$. The counting result $\boldsymbol{n}$ is a measure value. A sentence with **to count**$_{\mathrm{m}}$ translates as follows:

[7] I do not consider the following examples (which are brought up by Cécile Meier and an anonymous reviewer respectively): **John counted 10 ships when the lightning struck** or **John was (mis)counted as a supporter of the plan**. I leave the question open of how to incorporate these readings in my analysis and whether we should posit another ambiguity of **to count**.

(45) $|\textbf{John counted}_{\text{m}}\ \textbf{ten ships}| \equiv \mathbf{C}_i(\varepsilon, \textbf{john}', y_0)\ \&\ \mathbf{f}(\varepsilon) = \textbf{10 ships}'$,

where y_0 is a context-given group of objects to be counted and **10 ships'** is a counting result. In the implicit object y_0 I see another analogy to the measure verbs. In **John measured 3 meters** we have to assume an implicit object - a context-given object to be measured, too.

Accordingly to (45), the contribution of **to count**$_{\text{m}}$ is:

(46) $(\lambda v.(\lambda x.(\lambda \varepsilon.\mathbf{C}_i(\varepsilon, x, y_0)\ \&\ \mathbf{f}(\varepsilon) = v)))$,

where v is a measure-value variable. Because the domain of v contains only measure values, the ungrammaticality of sentences like (47) is predicted:

(47) **#John counted**$_{\text{m}}$ **ships/the ships/most ships/every ship.**

We have already seen that meausure phrases don't allow all determiners.

6.2 Semantic Solution

If we want to analyse the sortal-intensional reading as a genuine reading, we could assume that this reading is the underlying reading of the measure verb **to count**$_{\text{m}}$ and that the number-intensional reading – which is the extensional one – is derived from the intensional underlying reading (similar to the classical intensional verbs as **to look for**).

To satisfy the new assumption, i.e. the intensionality of the measure phrase object, we have to adjust function **f** introduced above. I call the new result function $\mathbf{f^i}$, with **i** standing for intensionality. The elements of the domain of the result function $\mathbf{f^i}$ are pairs consisting of an act of counting ε and a plural individual y_0, which is the objects to be counted. The range is a set of pairs consisting of the largest number n reached while counting and a property P of type $(s(et))$:

(48) $\mathbf{f^i}(\varepsilon, y_0) = (n, P)$ iff.
$[\![n]\!] = max\ m$, so that the following holds for m:
$(\exists z)[(\forall w \in \text{DOX}_{w_o, agent(\varepsilon)})[z \in [\![\mathbf{P}]\!](w) \cap [\![y_0]\!]]]$ and the agent of ε assings m to z in ε

The condition $z \in [\![\mathbf{P}]\!](w) \cap [\![y_0]\!]$ under the universal quantifier over possible worlds guarantees that not all objects of the context-given group have to be counted. With this condition we express the aspect of the meaning that the objects to be counted first have to be identified as such. Given a situation such that there are 20 birds in the pond – some of them ducks and some of them geese, (49) says that John had identified only 10 of them as ducks:

(49) **John counted**$_{\text{m}}$ **10 ducks.**

I think it is controversial, whether this is a semantic or pragmatic aspect of meaning. One can argue that picking out the objects to be counted takes place before counting. In this case we have to adjust the interpretation in (48) as follows:

(50) $\mathbf{f^i}(\varepsilon, \boldsymbol{y_0}) = (\boldsymbol{n}, \boldsymbol{P})$ gdw.
$[\![\boldsymbol{n}]\!] = max\ m$, so that the following holds for m:
$(\exists z)[(\forall w \in \mathrm{DOX}_{w_o,agent(\varepsilon)})[z \in [\![\mathbf{P}]\!](w) \wedge [\![\mathbf{P}]\!](w) = [\![\boldsymbol{y_0}]\!]]]$ and the agent of ε assings m to z in ε

I think both possibilities are conceivable.

In (51) we see the type logical translation of a sentence containing **to count**$_{\mathrm{m}}$ in the sortal-intensional reading:

(51) $|\mathbf{John\ counted_m\ 10\ ducks}|$
$\equiv \mathbf{C}_{\boldsymbol{i}}(\boldsymbol{\varepsilon}, \mathbf{john}\prime, \boldsymbol{y_0})\ \&\ \mathbf{f^i}(\boldsymbol{\varepsilon}, \boldsymbol{y_0}) = (\mathbf{10}, \lambda \boldsymbol{j}.\ \mathbf{duck}\prime_{\boldsymbol{i}})$,

To obtain the extensional reading, I will use the framework of Bäuerle (1983). It is well known that scope mechanisms can not always explain the intensional reading of indefinite NPs. Bäuerles framework deals with intensionality of indefinite NPs without reducing it to scope interaction. The main idea is that an NP can be interpreted relative to the evaluation situation or to the context. In (51), the situation variable $\boldsymbol{i}$, to which the constant $\mathbf{duck}\prime$ is applied, is not bound by the λ-prefix. For the intensional reading to be obtained, the condition must be fulfilled that $[\![\boldsymbol{j}]\!] = [\![\boldsymbol{i}]\!]$, otherwise extensional reading results.

The contribution of the verb is:

(52) $|\mathbf{count_m}| = (\lambda \boldsymbol{V}.(\lambda \boldsymbol{x}.(\lambda \boldsymbol{\varepsilon}.\mathbf{C}_{\boldsymbol{i}}(\boldsymbol{\varepsilon}, \boldsymbol{x}, \boldsymbol{y_0})\ \&\ \mathbf{f^i}(\boldsymbol{\varepsilon}, \boldsymbol{y_0}) = \boldsymbol{V})))$,

where the measure-value variable $\boldsymbol{V}$ is a pair consisting of a number and a property. Unlike the extensional measure phrase of the pragmatic solution, the measure phrase of this solution has an intensional component, a property.

7 A Potential Counterexample?

A potential problem for my analysis is presented by the following example:[8]

(53) **John (mis)counted 10 dogs as 6 foxes and 3 wolves.**

It seems that an extensional object (**10 dogs**) and a measure-phrase object (the number- and sortal-intensional **6 foxes and 3 wolves**) can appear with one occurrence of the surface **to count**. How could this be explained within my analysis of the surface **to count** as **to count**$_{\mathrm{t}}$ and **to count**$_{\mathrm{m}}$?

I think, the example (53) is not problematic for my analysis. There are reasons to assume that the **as**-phrase is *not* a measure phrase. Firstly, this position can be occupied by a non-measure phrase:

(54) **John (mis)counted dogs as foxes and wolves.**

Secondly, there are transitiv extensional verbs which also allow the same construction:

[8] I thank an anonymous reviewer for bringing these examples to my attention.

(55) **John bought apples as pears.**

It is obvious that the **as**-phrases in the above examples are not measure phrases.

I think, the feeling that the **as**-phrase in (53) is interpreted as a measure phrase (or a counting result), is again due to the fact that, by implicature (or default), the actual number of the objects counted is mistaken for the result of counting. As I already argued in section 6.1, the counting person assumes that the actual number of objects corresponds to the counting result. In other words, I take the **as**-phrase as relating to a property of objects (and their number) based on the counting result - but not the counting result itself. The exact analysis of such **as**-phrases should be carried out within a general theory of **as**-phrases.

8 Conclusion

In this paper I argued that the ambiguity between the extensional and non-extensional readings of **to count** should be analyzed as a case of polysemy between the transparent **to count**$_{\mathrm{t}}$ and the measure verb **to count**$_{\mathrm{m}}$, and not as a case of intensionality. At least in one of two so-called intensional readings I assume that **to count** is extensional. It is the measure-phrase object, which is responsible for the peculiar behavior of the verb, and not its intensionality.

References

1. Bäuerle, F.: Pragmatisch-semantische Aspekte der NP-Interpretation. In Faust, M., Harweg, R., Lehfeldt, W., and Wienold, G., editors, *Allgemeine Sprachwissenschaft, Sprachtypologie und Textlinguistik*, pages 121 – 132. Gunter Narr Verlag Tübingen. (1983)
2. Eschenbach, C.: *Zählangaben —- Maßangaben: Bedeutung und konzeptuelle Interpretation von Numeralia*. Deutscher Universitätsverlag. (1995)
3. Moltmann, F.: Intensional verbs and quantifiers. *Natural Language Semantics*, 5:1-52. (1997)
4. Moltmann, F.: *Abstract Objects and the Semantics of Natural Language*. Oxford University Press. (2013)
5. Montague, R.: On the nature of certain philosophical entities. In Thomason, R., editor, *Formal Philosophy. Selected Papers of Richard Montague*, pages 148 – 187. New Haven and London, Yale University Press. (1974)
6. Thomason, R. H.: Home is where the heart is. In French, P. A., Theodore E. Uehling, J., and Wettstein, H. K., editors, *Contemporary Perspectives in the Philosophy of Language*, pages 209 – 219. University of Minnesota press. (1979)
7. Wiese, H.: Semantische und konzeptuelle Strukturen von Numeralkonstruktionen. *Zeitschrift für Sprachwissenschaft*, 14/2:181 – 235. (1995)
8. Zifonun, G., Hoffmann, L., and Strecker, B.: *Grammatik der deutschen Sprache*, volume 3. de Gruyter. (1998)
9. Zimmermann, T.E.: Unspecificity and intensionality. In Féry, C. and Sternefeld, W., editors, *Audiatur Vox Sapientiae*, pages 514 – 533. Berlin. (2001)
10. Zimmermann, T.E.: Monotonicity in opaque verbs. Linguistics and Philosophy, 29:715 – 761. (2006)

A D-type Theory Solution to the Proportion Problem

Andreas Walker

University of Konstanz*

Abstract There are currently two major theories offering an analysis of donkey sentences: Dynamic Semantics (e.g. Kamp 1981, Heim 1982) and D-type theory (e.g. Heim 1990, Elbourne 2005). In their standard implementation, both theories generate so-called strong symmetric readings for donkey sentences. However, the literature also notes asymmetric readings of donkey sentences that have *usually* as their Q-adverb (the 'proportion problem'). Recently, Chen (2012) has argued that Dynamic Semantics have an advantage over D-type theory as they can solve the proportion problem by giving up unselective binding; in contrast, the most recent proposal for a D-type theory, Elbourne (2005), offers no solution to the problem. In this paper, I argue that Elbourne's system can straightforwardly be extended to solve the proportion problem. This will also provide a natural solution to the problem of weak readings, which allows us to reconsider the relationship between the weak/strong and the symmetric/asymmetric distinction.

1 The problem

In the analysis of donkey sentences, both dynamic semantics approaches (e.g. Kamp 1981, Heim 1982) and D-type theory (e.g. Heim 1990, Elbourne 2005) agree on generating the reading in (2) for the sentence in (1):

(1) If a farmer owns a donkey, he beats it.

(2) $\forall x \forall y$ [farmer(x) $\wedge$ donkey(y) $\wedge$ own(x,y) $\rightarrow$ beat(x,y)]

Dynamic approaches achieve this by having the covert Q-adverb *always* unselectively bind and universally quantify over both variables introduced by the two indefinites. D-type theories arrive at the same reading by having *always* quantify over minimal situations containing one farmer and one donkey each. Both approaches predict (4) to be true in the scenario described in (3):

(3) *There are three farmers. Two of them are poor and only own one donkey each. One farmer is very rich and owns fifty donkeys.*

(4) If a farmer owns a donkey, he is usually rich.

* The author would like to thank Maribel Romero, Irene Heim and three anonymous reviewers for their comments on this paper.

Intuitively, we judge the sentence in (4) to be false, which is at odds with the prediction. Both theories quantify over what amounts to farmer-donkey pairs. As the majority of these pairs is such that the farmer in them is rich (fifty cases compared to two), they arrive at the wrong prediction. What we actually seem to be quantifying over is just farmers. This problem is known as the proportion problem. The reading generated by the standard theories is known as the symmetric reading, whereas the reading observed in (4) is known as the asymmetric reading.

2 A dynamic solution

While both theories do not solve the proportion problem in their early standard implementations, there is a number of proposals in the dynamic literature (e.g. Dekker 1993, van Rooij 2006, Chen 2012, amongst others) that all involve giving up the idea of unselective binding. For the purpose of exposition, we will briefly sketch Dekker's version of the proposal here. This will only serve to show that Dynamic Semantics can in fact solve the problem. A detailed comparison between current dynamic approaches and D-type theory is not possible within the limits of this paper.

In his dissertation, Dekker (1993) develops EDPL, an update semantics based on Groenendijk and Stokhof's (1991) Dynamic Predicate Logic and Veltman's (1996) Update Semantics. EDPL handles unselectively quantifying adverbs by letting them quantify over output assignments. *Always* returns those assignments from an information state s, which, if they verify s updated with the restrictor, also verify s updated with the restrictor and the nuclear scope. That is, in (1), it returns those assignments which verify that for any farmer-donkey pair for which the owning-relation holds, the beating-relation also holds. The symmetric interpretation of *usually* works in parallel, returning those assignments in which for most farmer-donkey pairs for which the owning-relation holds, the beating relation also holds. The general update rule given for symmetric adverbs of quantification is below, where [A] is supposed to be read as the set-theoretic interpretation of the adverb in question:

(5) $s[A(\varphi)(\psi)] = \{i \in s | [A](\{j | i \leq j \ \& \ j \in s[\varphi]\})(\{j | j \in s[\varphi][\psi]\})\}$

In order to account for asymmetric readings, Dekker assumes that the adverbs come with a set of selection indices $\mathbf{X}$ that determine what the adverb quantifies over. The adverb then doesn't quantify over assignments, but over equivalence classes of assignments that agree in the specified variables. The update rule then looks as follows:

(6) If $X \subseteq (D(s[\varphi]) - D(s))$,
$s[A(\varphi)(\psi)] = \{i \in s | [A](\{j | i \leq_X j \ \& \ j \in s[\varphi]\})(\{j | j \in s[\varphi][\psi]\})\}$

An asymmetric reading of (1) that quantifies over farmers then simply checks whether all possible extensions of the assignment with individuals for y that

verify that y owns a donkey are also assignments that verify that y beats a donkey. That is, the number of donkeys beaten is irrelevant to the evaluation and every farmer is just counted once.

The exact implementation of the idea of giving up unselective binding does not matter for our purposes here. The sketch above only serves to illustrate that a solution is possible at all within dynamic semantics. Chen (2012) argues that this is one of the points where dynamic semantics approaches offer more flexibility in dealing with the problems that arise with donkey sentences. However, as we are going to show in the next section, a solution can also be implemented in D-type theory.

3 A D-type theory solution

Our analysis of asymmetric *usually* is a straightforward extension of the system presented in detail by Elbourne (2005). We begin by recalling the lexical entry that Elbourne gives for *always*. We will then conservatively extend this to an Elbournian *usually*, and finally modify this lexical entry so that it can deal with asymmetric readings.

3.1 Elbourne's analysis of *always*

Elbourne (2005) provides a D-type analysis based on Kratzer's (1989) situation semantics. Following Postal (1966), donkey pronouns are analysed as definite descriptions with NP-deletion. That is, the donkey sentence in (7) actually has the syntactic form in (8). As the pronoun itself is identical to the definite article, this is equivalent to a sentence of the form (9).

(7) If a farmer owns a donkey, he beats it.

(8) If a farmer owns a donkey, he ~~farmer~~ beats it ~~donkey~~.

(9) If a farmer owns a donkey, the farmer beats the donkey.

The definite article's uniqueness presuppositions are satisfied by evaluating it in minimal situations containing only unique individuals. Indefinite determiners make existential statements pertaining to those minimal situations, see (10), and those minimal situations are quantified over by the adverb, see (11):

(10) $[\![a]\!]^g = \lambda f.\lambda g.\lambda s.$ there is an individual x and a situation s' such that s' is a minimal situation such that s' $\leq$ s and f(λs.x)(s') = 1, such that there is a situation s" such that s" $\leq$ s and s" is a minimal situation such that s' $\leq$ s" and g(λs.x)(s") = 1

(11) $[\![always]\!]^g = \lambda p.\lambda q.\lambda s.$ for every minimal situation s' such that s' $\leq$ s and p(s') = 1, there is a situation s" such that s" $\leq$ s and s" is a minimal situation such that s' $\leq$ s" and q(s") = 1

For the sentence in (12), Elbourne assumes the LF in (13) and computes the truth conditions in (14)[1]:

(12) If a farmer owns a donkey, he beats it.

(13) [[always [if [[a farmer] [λ6 [[a donkey] [λ2 [t6 owns t2]]]]]]] [[he farmer] beats [it donkey]]]

(14) λs1. for every minimal situation s4 such that
s4 ≤ s1 and there is an individual y and a situation s7 such that s7 is a minimal situation such that s7 ≤ s4 and y is a man in s7, such that there is a situation s9 such that s9 ≤ s4 and s9 is a minimal situation such that s7 ≤ s9 and there is an individual x and a situation s2 such that s2 is a minimal situation such that s2 ≤ s9 and x is a donkey in s2, such that there is a situation s3 such that s3 ≤ s9 and s3 is a minimal situation such that s2 ≤ s3 and y owns x in s3,
there is a situation s5 such that
s5 ≤ s1 and s5 is a minimal situation such that s4 ≤ s5 and ιx x is a man in s5 beats ιx x is a donkey in s5

As it is very hard to parse these truth conditions, Elbourne provides an illustration that is reproduced in (Fig. 1). From this we can see that the relevant situations for our purposes are s4 and s5. We give a rough paraphrase of the truth conditions in (15), leaving out those parts of the structure that are not directly relevant.

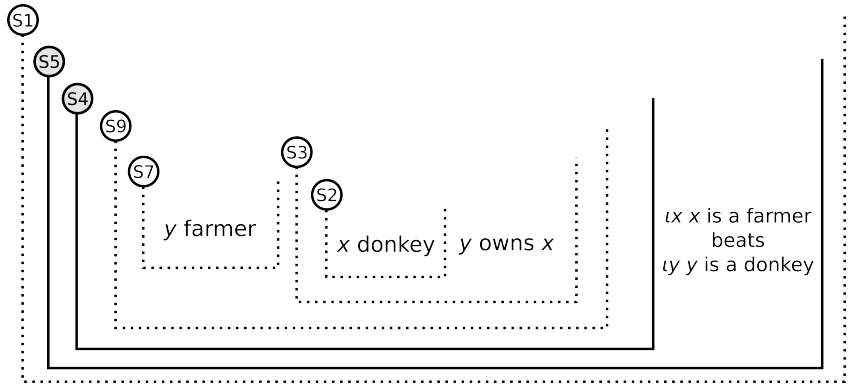


Figure 1: Nesting of situations in Elbourne (2005: 52)

(15) For every s4 in the set S4 of minimal situations containing one farmer owning one donkey, there is an extension of s4 in the set S5 of minimal situations containing one farmer owning and beating one donkey.

[1] For a detailed computation and additional explanations of the (rather lengthy) truth conditions, the reader is referred to Elbourne's (2005: 249) Appendix B.1.

3.2 An Elbournian analysis of symmetric *usually*

We are now concerned with a variant of (12) that has *usually* as its Q-adverb:

(16) If a man owns a donkey, he usually beats it.

Elbourne (2005) does not provide a lexical entry for *usually*. However, from what we have seen above, we can reconstruct one that should be in his spirit. The expected truth conditions of the sentence in (16) are, in abbreviated form:

(17) $|\text{S5}| > \frac{1}{2}\ |\text{S4}|$

That is, we expect more than half of the minimal situations containing one farmer owning one donkey to have an extension containing one farmer owning and beating one donkey. A lexical entry to derive these truth conditions is (18):

(18) $[\![usually]\!]^g = \lambda p.\lambda q.\lambda s.$ | { s": s' is a minimal situation such that s' $\leq$ s and p(s') = 1 and s" is a situation such that s" $\leq$ s and s" is a minimal situation such that s' $\leq$ s" and q(s") = 1 } |
$> \frac{1}{2}$ | { s': s' is a minimal situation such that s' $\leq$ s and p(s') = 1 } |

This lexical entry compares the cardinality of the set of minimal farmer-donkey situations to the cardinality of the set of their extensions. This gives us the symmetric reading in which farmer-donkey pairs are counted.

3.3 Making *usually* asymmetric

In order to derive asymmetric readings, we could pursue two directions: We either modify the lexical entries for the indefinite article, so that it will return situations containing one farmer and some, possibly all of his donkeys, or we modify the lexical entry for the Q-adverb in (18) so that it derives an asymmetric reading from minimal farmer-donkey pair situations. The first approach, which might be traced back to Berman (1987), would require us to find a way of generating situations of just the right size, possibly by using Kratzer's (2012) notion of exemplification and some way of shifting the interpretation of the indefinite towards a plural or mass reading. Even if this could be solved in a convincing manner, it would immediately present us with a new problem: As minimal situations were introduced to D-type theory in the first place to deal with the problem of the definite article's uniqueness presuppositions, we would reintroduce this problem into D-type theory. Since we want to avoid this, we will pursue the second approach. Our basic idea is that *usually* does some additional work on the two sets constructed in the lexical entries in (11) and (18). Instead of simply comparing their cardinality, it will first apply a function $\textsc{Fuse}_K$ to both of them. The comparison is then run on the output of this function. That is, the truth conditions we assume are, abbreviated:

(19) $|\textsc{Fuse}_K(\text{S5})| > \frac{1}{2}\ |\textsc{Fuse}_K(\text{S4})|$

And the lexical entry for *usually* is only minimally modified to read:

(20) $[\![usually]\!]^g = \lambda p.\lambda q.\lambda s.$ | FUSE$_K$({ s": s' is a minimal situation such that s' $\leq$ s and p(s') = 1 and s" is a situation such that s" $\leq$ s and s" is a minimal situation such that s' $\leq$ s" and q(s") = 1 }) | $> \frac{1}{2}$ | FUSE$_K$({ s': s' is a minimal situation such that s' $\leq$ s and p(s') = 1 }) |

(21) *to be revised*
FUSE$_K$ = λS. { the minimal situation s such that
$\forall s' \in S[x \leq s' \rightarrow s' \leq s] : x \in K\}$
where **K** is a contextually given set of individuals[2]

That is, FUSE$_K$ merges all those situations that share an individual from the sorting key **K**. A short derivation for the example in (22), with the LF in (23), will show that this indeed produces the desired reading.

(22) If a farmer owns a donkey, he usually beats it.

(23) [[usually [if [[a farmer] [λ6 [[a donkey] [λ2 [t6 owns t2]]]]]]] [[he farmer] beats [it donkey]]]

We judge (22) false in the scenario (24), while it comes out as true on the standard analysis:

(24) *There are three farmers. Farmer 1 and Farmer 2 have one donkey each which they do not beat, but Farmer 3 beats all of his three donkeys.*

For reasons of space, we ask the reader to consult Elbourne's (2005: 249) Appendix B.1 for the first part of the derivation. The only point where we depart from Elbourne is the lexical entry of the adverb. Inserting it into the derivation, a few instances of functional application get us to the following point:

(25) λs. |FUSE$_K$({ s5: s4 is a minimal situation such that
s4 $\leq$ s and there is an individual y and a situation s7 such that s7 is a minimal situation such that s7 $\leq$ s4 and y is a man in s7, such that there is a situation s9 such that s9 $\leq$ s4 and s9 is a minimal situation such that s7 $\leq$ s9 and there is an individual x and a situation s2 such that s2 is a minimal situation such that s2 $\leq$ s9 and x is a donkey in s2, such that there is a situation s3 such that s3 $\leq$ s9 and s3 is a minimal situation such that s2 $\leq$ s3 and y owns x in s3
and s5 is a situation such that
s5 $\leq$ s and s5 is a minimal situation such that s4 $\leq$ s5 and ι x x is a man in s5 beats in s5 ι x x is a donkey in s5})|
$> \frac{1}{2}$ |FUSE$_K$({s4: s4 is a minimal situation such that
s4 $\leq$ s and there is an individual y and a situation s7 such that s7 is a minimal situation such that s7 $\leq$ s4 and y is a man in s7, such that there is a situation s9 such that s9 $\leq$ s4 and s9 is a minimal situation

[2] We will revise this assumption later, which will slightly complicate FUSE$_K$.

such that s7 $\leq$ s9 and there is an individual x and a situation s2 such that s2 is a minimal situation such that s2 $\leq$ s9 and x is a donkey in s2, such that there is a situation s3 such that s3 $\leq$ s9 and s3 is a minimal situation such that s2 $\leq$ s3 and y owns x in s3})|

Again, it is useful to roughly paraphrase this to see what is going on:

(26) λs. |FUSE$_K$(**the extended situations S5**)|
$> \frac{1}{2}$ |FUSE$_K$(**the minimal situations S4**)|

In our scenario (24), the set of minimal situations S4 will consist of five, and the set of extended situations S5 of three situations, as shown in (27) and (28) with each situation in square brackets:

(27) S4 = {[f1 owns d1], [f2 owns d2], [f3 owns d3], [f3 owns d4], [f3 owns d5]}

(28) S5 = {[f3 owns & beats d3], [f3 owns & beats d4], [f3 owns & beats d5]}

Without application of FUSE$_K$ the sentence would come out as true, because |S5| $> \frac{1}{2}$ |S4| (that is, 3 > 2.5). However, under the assumption that the context is such that **K** = {f1, f2, f3} (i.e. the set of farmers), applying FUSE$_K$ will result in the sets (29) and (30):

(29) FUSE$_K$(S4) = {[f1 owns d1], [f2 owns d2], [f3 owns d3 & d4 & d5]}

(30) FUSE$_K$(S5) = {[f3 owns & beats d3 & d4 & d5]}

Now, the sentence comes out as false, as we intended (1 $\not>$ 1.5).

Our use of FUSE$_K$ is based on the same idea as Beck and Sharvit's (2002) analysis of quantificational variability effects in questions: In Beck and Sharvit, the relevant adverb embeds a question that can be partitioned into more fine-grained subquestions. In our case, we are going in the reverse direction: the set of minimal situations is maximally fine-grained, and to arrive at the desired domain of quantification for the adverb, we fuse together some of the situations. This 'fusing together' is the task of FUSE$_K$. The appropriate way of fusing situations in a given case is based on a contextually supplied sorting key **K**.

Obviously, the idea of **K** is also related to the dynamic semantics **X**. However, while **X** directly specifies variables to be quantified over, we assume that **K** instead specifies a list of individuals, e.g. the set of farmers or the set of donkeys. Of course this can also be generalized to cover both symmetric readings and asymmetric readings with an arbitrary number of variables. In order to do this, we let **K** be a set of n-tuples of individuals:

(31) FUSE$_K$ = λS. { the minimal situation s such that
$\forall s' \in S, y \in X[y \leq s' \rightarrow s' \leq s] : X \in K$}
where **K** is a contextually given set of n-tuples of individuals

In the asymmetric cases where we quantify only over farmers or donkeys, **K** contains singletons of individuals and FUSE$_K$ produces the same results as above.

In the symmetric cases, **K** contains farmer-donkey pairs and $\textsc{Fuse}_K$ simply returns the original set of minimal situations. For asymmetric cases with up to *n* variables, **K** will accordingly contain tuples of *n-1* elements.

To illustrate this, consider again sentence (22) in the scenario (24). Different assumptions about the content of **K** will then yield the different readings:

(32) **K** = {<f1, d1>, <f2, d2>, <f3, d3>, <f3, d4>, <f3, d5>}
symmetric reading

(33) **K** = {<f1>, <f2>, <f3>}
asymmetric reading, counting farmers

(34) **K** = {<d1>, <d2>, <d3>, <d4>, <d5>}
asymmetric reading, counting donkeys

4 Extending the analysis to *always*: weak readings

As we have seen above, our proposal solves the proportion problem for *usually*. But for *always*, there is an additional reading that deviates from the standard symmetric reading of donkey sentences. So-called weak readings are known to be available in sentences like (35):

(35) If a man has a dime, he (always) throws it into the meter.

Here, we do not require a man to throw all of his dimes into the meter. Rather, throwing one dime seems to suffice. As it turns out, applying our analysis of *usually* to *always* will can easily obtain this reading:

(36) $[\![always]\!]^g = \lambda p.\lambda q.\lambda s.$ | $\textsc{Fuse}_K$({ s": s' is a minimal situation such that s' ≤ s and p(s') = 1 and s" is a situation such that s" ≤ s and s" is a minimal situation such that s' ≤ s" and q(s") = 1 }) | = | $\textsc{Fuse}_K$({ s': s' is a minimal situation such that s' ≤ s and p(s') = 1 }) |

This lexical entry utilizes $\textsc{Fuse}_K$ in exactly the same way as that for *usually*, with the only difference being the strength of the quantification. Instead of requiring the cardinality of $\textsc{Fuse}_K$(S4) to be more than half of the cardinality of $\textsc{Fuse}_K$(S5), we now require the cardinality of the two sets to be identical. The weak reading naturally falls out if **K** contains singletons of men, while the strong reading can still be generated by requiring **K** to contain men-dime pairs. Consider the following scenario:

(37) There are two men. Each man has two dimes and throws one of them into the meter.

In our scenario, the set of minimal situations S4 will consist of four, and the set of extended situations S5 of two situations, with each situation in square brackets:

(38) S4 = {[m1 has d1], [m1 has d2], [m2 has d3], [m2 has d4]}

(39) S5 = {[m1 has & throws d1], [m2 has & throws d3]}

Under the assumption that **K** = {<m1>, <m2>}, applying $\textsc{Fuse}_K$ will result in the sets (40) and (41):

(40) $\textsc{Fuse}_K$(S4) = {[m1 has d1 & d2], [m2 has d3 & d4]}

(41) $\textsc{Fuse}_K$(S5) = {[m1 has & throws d1], [m2 has & throws d3]}

Since the cardinality of both sets is the same (i.e. 2), the sentence comes out as true, as intended under the weak reading. In order to derive the strong reading, **K** would have to be {<m1, d1>, <m1, d2>, <m2, d3>, <m2, d4>}, parallel to the example in (32). Then, applying $\textsc{Fuse}_K$ would return the original sets in (38) and (39). Since their cardinality is not the same ($|\textsc{Fuse}_K(\text{S4})| = 4$, $|\textsc{Fuse}_K(\text{S5})| = 2$), the sentence would come out as false.

5 Back to *usually*

At first, the two distinctions weak/strong and symmetric/asymmetric look as if they are orthogonal to each other, yielding four possible readings between them: [i] weak-asymmetric, [ii] strong-asymmetric, [iii] weak-symmetric and [iv] strong-symmetric. However, the symmetric cases always come out the same. Since there is only one donkey in a minimal situation containing a farmer and a donkey, requiring the farmer to beat all of his donkeys in this situation is not different from requiring him to beat one of his donkeys in this situation. Thus, [iii] and [iv] are identical. In light of this we should think of the weak/strong distinction as one that applies to asymmetric readings only.

In the case of *usually*, we could in principle expect to see both readings [i] and [ii]. Our analysis, as sketched above, yields the weak-asymmetric reading [i]. If this is indeed the only reading we observe, we can collapse both distinctions into one, as is the case with *always*: there is a symmetric (strong) reading and an asymmetric (weak) reading. I am not currently aware of any data contradicting this prediction. One possible exception might be Kanazawa (1994) who claims a strong asymmetric reading for sentences of the following kind:

(42) Most farmers that own a donkey also own its offspring.

However, note that in all scenarios that make the sentence true on the strong reading, it is also true on the weak reading. There are indeed scenarios where the sentence seems false although it would come out as true on the weak asymmetric reading. But it might simply be the case that the asymmetric reading is not available at all in those cases. Judgements in proportional donkey sentences are notoriously difficult to obtain, and a definite verdict on this might require experimental research beyond the scope of this present paper. If sentences like (42) turn out to have a strong asymmetric reading, it is easy to accommodate this in our system. We simply have to allow for **K** to contain tuples that contain one farmer each together with the sum of all his donkeys. Consider the following case:

(43) There are two farmers who own two donkeys each. One of the farmers owns the offspring of one of his donkeys, the other farmer owns the offspring of both of his donkeys.[3]

Then we would judge the sentence in (42) true under the weak asymmetric reading[4] and the symmetric reading, but false under the strong asymmetric reading. The situations would look like this:

(44) S4 = {[f1 owns d1], [f1 owns d2], [f2 owns d3], [f2 owns d4]}

(45) S5 = {[f1 owns d1 & d1's offspring], [f2 owns d3 & d3's offspring], [f2 owns d4 & d4's offspring]}

And fusing them under the assumption that **K** has the form in (46) will yield the sets in (47) and (48):

(46) **K** = {<f1, ⊕{d1, d2}>, <f2, ⊕{d3, d4}>}
strong asymmetric reading (by farmers)

(47) $\textsc{Fuse}_K$(S4) = {[f1 owns d1 & d2], [f2 owns d3 & d4]}

(48) $\textsc{Fuse}_K$(S5) = {[f2 owns d3 & d4 & d3's offspring & d4's offspring]}

Then the sentence comes out false, because the cardinality of $\textsc{Fuse}_K$(S5) is exactly half the cardinality of $\textsc{Fuse}_K$(S4), and not less than it.

6 Conclusions

As our analysis has shown, D-type theory can be extended to account for the proportion problem while keeping the advantages of Elbourne's (2005) approach with respect to the uniqueness presuppositions of the definite article, with a solution for the problem of weak readings naturally following from this extension. While our approach relies on the context to supply a way of structuring the Q-adverb's domain of quantification, it does so in a fairly constrained way. In fact, the contextual variable **K** that we use has a clear counterpart in dynamic semantic theories: the analysis by Dekker (1993), for example, uses a set of variables **X** to decide what variables should be quantified over.

Both theories still need to be connected to features of information structure in a principled way. For example, we know that focus plays a role in determining which asymmetric readings are available, as illustrated in the examples below, quoted from Heim (1990):

(49) If a DRUMMER lives in an apartment complex, it is usually half empty.

(50) If a drummer lives in an APARTMENT COMPLEX, it is usually half empty.

[3] To avoid complications, assume that all the offspring the donkeys in this scenario have produced are mules, i.e. we don't have to count the offspring among the donkeys.

[4] This is slightly complicated by the fact that *most* comes with an implicature of 'not all'. We are ignoring this for now to keep the scenario simple.

However, there is no overall satisfying account of the connection between various surface phenomena and the available readings of donkey sentences. Generally, the exact interaction between information structure and the availability of donkey sentence readings remains in need of more empirical data.

One point where the two analyses possibly diverge is the flexibility provided by the respective variables. While **X** is very constrained in that it can only contain or exclude entire variables, **K** is rather flexible and could in principle contain very diverse tuples of individuals (e.g. pairs of a farmer and the sum of some, but not all, of his donkeys). In our cases, both amount to the same, but it will be necessary to empirically verify the available readings in order to see whether this added flexibility might prove to be necessary.

However, the flexibility that **K** provides also comes with a potential downside. In this paper, we assume that $\textsc{Fuse}_K$ simply checks whether an individual participates in a situation. This might become problematic in cases where one and the same individual participates in more than one role, such as (51):

(51) If a farmer hires a man to drive his tractor, he is usually rich.

Here, all male farmers potentially appear in situations both as an employer and as an employee. $\textsc{Fuse}_K$ will not distinguish between these roles, and subsequently the fused situations might not produce the correct domain of quantification. While I do not have a solution to this potential problem at present, there are several approaches that could be taken for solving it, such as giving more structure to the contents of **K**, for example by assuming that **K** is actually a set of situations rather than individuals, so that information about specific roles within a given situation can be specified. However, the nature of this problem suggests that it is a variant of the problem of indistinguishable participants. If that is the case, it applies to the D-type approach as a whole, and is not specific to my proposal here.

In the absence of additional data, it looks like dynamic semantics and D-type theory are at least back on the same level with respect to the proportion problem – both can account for it using roughly the same resources. Our analysis also straightforwardly extends to weak readings, which were not generated by Elbourne (2005). This also shows that D-type theory is less constrained by the minimality of situations than it might appear at first sight. The system becomes quite flexible if we equip it with a very simple mechanism for fusing situations, while keeping intact its ability to treat donkey pronouns as definite descriptions. What remains for future work is a detailed comparison of D-type theory and current approaches to dynamic semantics.

References

1. Beck, S., Sharvit, Y.: Pluralities of questions. Journal of Semantics 19, 105–157 (2002)
2. Berman, S. R.: Situation-based Semantics for Adverbs of Quantification. UMass Occasional Papers 12, 45–68 (1987)
3. Chen, H.-Y.: Donkey Pronouns. University of Texas at Austin Dissertation (2012)
4. Dekker, P.: Transsentential meditations. University of Amsterdam Dissertation (1993)
5. Elbourne, P.: Situations and Individuals. Cambridge, MA: MIT Press (2005)
6. Groenendijk, J., Stokhof, M.: Dynamic Predicate Logic. Linguistics and Philosophy 14, 39–100 (1991)
7. Heim, I.: The semantics of definite and indefinite noun phrases. Amherst: GLSA (1982)
8. Heim, I.: E-type pronouns and donkey anaphora. Linguistics and Philosophy 13, 137–177 (1990)
9. Kamp, H.: A theory of truth and semantic representation. In Portner, P. and Partee, B. (eds.): Formal Semantics. The Essential Readings. Oxford: Blackwell, 189–222 (1981)
10. Kanazawa, M.: Weak vs. strong readings of donkey sentences and monotonicity inference in a dynamic setting. Linguistics and Philosophy 17, 109–158 (1994)
11. Kratzer, A.: An investigation of the lumps of thought. Linguistics and Philosophy 12(5), 607–653 (1989)
12. Kratzer, A.: Modals and Conditionals. Oxford: Oxford University Press (2012)
13. Postal, P.: On so-called pronouns in English. In F. Dinneen, ed., Report on the 17th annual round table meeting on linguistics and language studies, 177–206 (1966).
14. van Rooij, R.: Free choice counterfactual donkeys. Journal of Semantics 23(4), 383–402 (2006)
15. Veltman, F.: Defaults in Update Semantics. Journal of Philosophical Logic 25(3), 221–261 (1996)

Analysis and Implementation of Focus and Inverse Scope by Delimited Continuation

Youyou Cong

Ochanomizu University,
2-1-1, Otsuka, Bunkyo-ku, Tokyo, Japan
so.yuyu@is.ocha.ac.jp

Abstract Focus and inverse scope are known as phenomena that require the context of a particular lexical item for the semantic representation of the whole sentence. The context of a given lexical item in a natural language sentence can be regarded as its continuation. Continuation is a notion in programming languages that represents the rest of the computation. In this paper, we will present an analysis and an implementation of focus and inverse scope by means of the control operators shift/reset (Danvy and Filinski 1990). We also discuss the interaction between these phenomena.

1 Introduction

Focus and inverse scope are phenomena that require the context of a particular lexical item for the semantic representation of the whole sentence. In generative grammar, one can obtain the context by "covert movements" (May 1977, Wagner 2006); by contrast, these phenomena are a major challenge for grammars that do not allow movements, such as categorial grammar.

Meanwhile, computer scientists have studied various applications of continuations. Continuations represent the rest of the computation. The context of a given lexical item in a natural language sentence can be regarded as its continuation. In fact, the notion of continuations has been applied to natural language semantics in the last decade (de Groote 2001, Barker 2002, Barker 2004, Bekki and Asai 2010). Currently, however, there seems to be no comprehensive theory that accounts for both focus and inverse scope using continuation. In particular, nested focus (Krifka 1991) and inverse scope construction involving more than two quantifiers pose a problem to previous theories.

In this paper, we propose an analysis and an implementation of focus and inverse scope by using the control operators `shift/reset` (Danvy and Filinski 1990). In our analysis, the meaning of sentences that contain focus or inverse scope can be calculated without movements or other mechanisms such as quantifier storage. In the rest of the paper, we will introduce the notion of continuations, discuss the previous work on focus and inverse scope in both formal semantics and computer science, and then present our approach. We implement our account in OchaCaml (Masuko and Asai 2011), a variant of a functional programming language called ML. The implementation allows the easy and reliable

testing of complex examples involving focus and quantification. We also discuss the interaction between focus and inverse scope.

2 Continuations

2.1 Continuations and CPS Transformation

Continuations represent the rest of the computation. For example, when you are calculating $(2 * 3)$ in the expression $1 + (2 * 3) - 4$, the current continuation is paraphrased as: "given the value of the current computation, add 1 to it and subtract 4 from the sum". In the lambda calculus, this is the function $\lambda x.\,(1 + x - 4)$.

Continuations are not explicit in programs written in direct style. Transforming the whole program into continuation-passing style (CPS) makes the continuation of each subterm visible. In programs written in CPS, each function receives an additional argument for its continuation, which represents what to do with the value of the function. A formal definition of Call-by-Value CPS transformation is the following (Plotkin 1975):

$$\begin{aligned} [\![x]\!] &= \lambda k.\, k\, x \\ [\![\lambda x.\, M]\!] &= \lambda k.\, k\, (\lambda x.\, [\![M]\!]) \\ [\![M\, N]\!] &= \lambda k.\, [\![M]\!]\, (\lambda m.\, [\![N]\!]\, (\lambda n.\, (m\, n)\, k)) \end{aligned}$$

In the definition above, $[\![\ ]\!]$ denotes the translation function, and k represents the continuation of the expression that one wants to transform into CPS. When the expression is a variable, no further computation is required, so it is simply passed to the current continuation. Note that $\lambda x.\, k\, x$ denotes $\lambda x.\, (k\, x)$. In the case of lambda abstraction, the body M is translated into CPS and its abstraction is passed to the continuation. The translation of function application may be helpful to see the behavior of CPS programs. First, the function M is computed, and its value is bound by the λ operator as m. Next, the argument N is processed in the same way, with its value n bound by the λ operator. Then, the application $m\, n$ is evaluated in the context k.

In CPS programs, we can specify the order of evaluation. If we adopt the definition above, the order of evaluation is left-to-right. If we transform the function application $M\, N$ into $\lambda k.\, [\![N]\!]\, (\lambda n.\, [\![M]\!]\, (\lambda m.\, (m\, n)\, k))$ (i.e., if we let the argument be evaluated first), the program will be evaluated right-to-left.

2.2 `shift/reset`

CPS transformation requires transforming the entire program. The control operators `shift/reset` (Danvy and Filinski 1990) enable us to handle continuations in direct style. The `shift` operator captures the current continuation, and the `reset` operator delimits the continuation captured by `shift`. Consider the computation below.

(1) `1 + reset (2 * shift k. k (k 3))`

The continuation `k` captured by the `shift` operator is the function $\lambda x.\,(2 * x)$. This function is applied twice to 3, and then the calculation of adding 1 outside the `reset` is executed. The result of (1) will be 13.

More formal definition of `shift/reset` operators can be defined in terms of CPS transformation.

$$[\![\texttt{shift}\, c.\, M]\!] = \lambda k.\, [\![M]\!][\lambda v.\, \lambda k'.\, k'\,(k\, v)/c]\,(\lambda x.\, x)$$
$$[\![\texttt{reset}\,(M)]\!] = \lambda k.\, k\,([\![M]\!]\,(\lambda x.\, x))$$

The `shift` operator binds the current continuation to the variable c, and evaluates the body M with an empty context. The notation $[\lambda v.\, \lambda k'.\, k'\,(k\, v)/c]$ means substituting $\lambda v.\, \lambda k'.\, k'\,(k\, v)$ for c. The `reset` operator evaluates the expression M with an empty context and passes its result to the context surrounding the `reset` clause.

3 Focus

3.1 Alternative Semantics

Focus has been explicated as the most important or new information in an utterance. In the conversation below, the new information for A is "Mary", which may be focused (we enclose the focused phrase with $[\ \ ]_{\mathrm{F}}$).

(2) A. Who does John love?
B. John loves $[\mathrm{Mary}]_{\mathrm{F}}$.

Rooth (1992, 1996) claimed that focus evokes those alternatives that are relevant for the interpretation of the focused expression. According to Rooth, the sentence (2B) gives rise to a set of propositions of the form "John loves x", where x is "Mary" or some other person relevant in the situation. This set is called the alternative set.

Now, let us consider the combination of focus and the adverb "only".

(3) John only loves $[\mathrm{Mary}]_{\mathrm{F}}$.

(3) means that there is no true proposition of the form "John loves x" except for "John loves Mary". In other words, among the propositions in the alternative set of (3), only the element "John loves Mary" is true. Rooth (1996) gives the semantic representation of "only" as follows.

(4) $\lambda C.\, \lambda p.\, \forall q\, [q \in C \wedge {}^{\vee}q \leftrightarrow (q = p)]$

Here, p is the proposition where the focused phrase is substituted for the focused position and C is the alternative set. In the case of (3), p is the proposition "John loves Mary" and C may be the set {John loves Mary, John loves Sue, John loves Alice}. ${}^{\vee}q$ denotes that q is true at a given possible world. (4) means that for every proposition q in C, q is true iff $q = p$.

3.2 Bekki and Asai (2010)

To represent the meaning of a sentence that contains a focus, we need a proposition in which the focused phrase is abstracted. Such a proposition can be regarded as the continuation of the focused phrase. Bekki and Asai (2010) uses `shift` to abstract focused phrases, and `reset` to delimit the continuations. The semantic representation of "only" and its focus is defined as follows.

$$[M]_{\mathrm{F}} = \mathit{shift}\, k.\, \forall x\, (k\, x \leftrightarrow x = M)$$
$$\mathit{only}\, (\varphi) = \mathit{reset}\, (\varphi)$$

The continuation k captured by the `shift` operator is the proposition in which the focused phrase is abstracted as a variable. The definition above means that for every x, the proposition in which x is substituted for the variable is true iff x is the focused phrase.

Following this definition, the sentence (3) can be represented as follows (where j and s are the denotations of John and Mary, respectively):

(5) $\llbracket \mathit{only}\, (\mathit{love}\, (j,\, [m]_{\mathrm{F}})) \rrbracket$
$= \llbracket \mathit{reset}\, (\mathit{love}\, (j,\, (\mathit{shift}\, k.\, \forall x\, (k\, x \leftrightarrow x = m)))) \rrbracket$
$= \forall x\, (\mathit{love}\, (j,\, x) \leftrightarrow x = m)$

The continuation k captured by the `shift` operator is the function $\lambda y.\, \mathit{love}\, (j,\, y)$. The reduced representation means that for all x, John loves x iff $x =$ Mary. This is equivalent to the conjunction of "John loves Mary" and "John loves no one other than Mary".

3.3 Nested Focus

A single focus may be associated with two distinct adverbs. Such a focus can be represented by means of a nested focus structure (Krifka 1991). Consider the sentence (6b), which presupposes that there is some x other than wine such that John only drank x in the past. We enclose the focus of "only" with $[\ ]_{\mathrm{F_o}}$, and the focus of "also" with $[\ ]_{\mathrm{F_a}}$.

(6) a. Last month John only drank $[\text{beer}]_{\mathrm{F_o}}$.
b. He has also only drunk $[[\text{wine}]_{\mathrm{F_a}}]_{\mathrm{F_o}}$.

In (6b), "wine" has a nested focus construction, where the outer focus is associated with "only" and the inner focus with "also". Rooth (1996) shows the derivation of this nested focus.

(7) also $[_{\mathrm{S}}\ [\text{wine}]_{\mathrm{F_a}}\ \lambda e_2\ [_{\mathrm{S}}$ have $[_{\mathrm{S}}$ only $[_{\mathrm{S}}\ [e_2]_{\mathrm{F_o}}\ \lambda e_1\ [_{\mathrm{S}}$ He drunk $e_1]]]]]$

Intuitively, we can represent a nested focus by using nested `shift` operators. Here we represent the presupposition brought by "also" as $\exists y\, (k\, y \wedge \neg(y = M))$, where M is the phrase focused by "also" and k is its continuation[1]. This means

[1] The meaning of the focus associated with "also" is represented as $\exists y\, (k\, y \wedge \neg(y = M)) \wedge (k\, M)$. For the sake of simplicity, here we omit the assertive part $k\, M$.

that there exists some y such that the proposition where the focused phrase is replaced with y holds and y is distinct from M. Then (6b) is analyzed as follows:

(8) $[\![reset\,(reset\,(drink\,(j,\ shift\,k_o.\forall x\,(k_o\,x \leftrightarrow x = shift\,k_a.\exists y\,(k_a\,y \wedge \neg(y = w))))))]\!]$
$= \exists y\,(\forall x\,(drink\,(j,\ x) \leftrightarrow x = y) \wedge \neg(y = w))$

Here, k_o is the proposition "John drank x", and k_a is "John only drank $[x]_{\mathrm{F_o}}$". In this case the representation can be reduced as desired. However, things are different in the sentence (9), whose expected representation is (10).

(9) Sue has also thought that John only loves $[[\mathrm{Mary}]_{\mathrm{F_a}}]_{\mathrm{F_o}}$.

(10) $\exists y\,(think\,(s,\ \forall x\,(love\,(j,\ x) \leftrightarrow x = y)) \wedge \neg(y = m))$

This sentence has a reading which presupposes that there is some x such that Sue thought that John only loves x, where x is distinct from Mary. In this case, the scope of "only" is "John loves x", and the scope of "also" is "Sue has thought that John only loves $[x]_{\mathrm{F_o}}$". (11) is the representation with `shift/reset`, which is reduced in the following way.

(11) $[\![reset\,(think\,(s,\ reset\,(love\,(j,\ shift\,k_o.\forall x\,(k_o\,x \leftrightarrow x = shift\,k_a.\exists y\,(k_a\,y \wedge \neg(y = m)))))))]\!]$
$= think\,(s,\ \exists y\,(\forall x\,(love\,(j,\ x) \leftrightarrow x = y) \wedge \neg(y = m)))$

In the reduced representation, the existential quantifier takes scope within the predicate "think". This is because the `shift` operator captures the continuation that is delimited by the innermost `reset` operator. In this case, the outer `shift` operator is evaluated first, with its continuation $love\,(j,\ x)$ bound to k_o. What should be noted here is that the body of the `shift` operator is evaluated with an empty context, as can be seen in the definition given in Section 2. This functions in the same way as the `reset` operator, so the inner `shift` operator cannot capture the computation outside the body of the first `shift` operator. Thus, the continuation k_a becomes $\forall x\,(love\,(j,\ x) \leftrightarrow x = y)$. As a consequence, the existential quantifier only takes scope over this expression. We will show how to deal with such a case using `shift/reset` in Section 5.

4 Inverse Scope

When a sentence has two quantifiers, it may have two readings: surface scope reading and inverse scope reading. Consider the sentence (12). (12a) is the surface scope reading and (12b) is the inverse scope reading.

(12) Some woman loves every man.
  a. $\exists x\,(woman\,(x) \wedge \forall y\,(man\,(y) \rightarrow love\,(x,\ y)))$
  b. $\forall y\,(man\,(y) \rightarrow \exists x\,(woman\,(x) \wedge love\,(x,\ y)))$

(12a) represents the reading that there is some specific woman who loves every man. In contrast, (12b) represents the reading that for each man y, there is some woman who loves y. In the representation (12a), the scopes of the two quantifiers have the same order as the surface structure, while in (12b) the scope order is inverted.

4.1 Quantifier Raising

May (1977) suggests that all the possible readings of a sentence containing more than two quantifiers can be represented with Logical Form (LF) by applying Quantifier Raising (QR) to each quantified NP. QR is an operation applied while transforming the Surface Structure (SS) of a sentence to its LF. When we apply QR to some quantified NP, it moves to the adjunct of the minimal S-node that contains it, with its trace left in the scope. The LF that corresponds to the surface scope reading of (12) can be derived by applying QR to the subject NP first and then to the object NP; and if we reverse this order, the inverse scope reading is derived. The following LFs represent the two readings.

(13) a. $[_{\mathrm{S}}\,[_{\mathrm{NP}}\,\text{some woman}]_2\;[_{\mathrm{S}}\,[_{\mathrm{NP}}\,\text{every man}]_3\;[_{\mathrm{S}}\,e_2\;[_{\mathrm{VP}}\,\text{loves}\;e_3]]]]$

b. $[_{\mathrm{S}}\,[_{\mathrm{NP}}\,\text{every man}]_3\;[_{\mathrm{S}}\,[_{\mathrm{NP}}\,\text{some woman}]_2\;[_{\mathrm{S}}\,e_2\;[_{\mathrm{VP}}\,\text{loves}\;e_3]]]]$

In (13a), the second S-node is the context of "some woman", and the third S-node is the context of "every man". Similarly, in (13b), the second S-node and the third S-node represent the context of "every man" and "some woman", respectively. Thus, applying QR makes the context of each quantified NP explicit. One can see that in LF, the quantifier that takes the broader scope contains other quantifiers in its context.

4.2 Barker (2002)

Barker (2002) presents an analysis of quantification in the framework called "continuized semantics", where each lexical item is transformed into CPS so as to make its continuation explicit. According to Barker, scope order depends on the order of evaluation. For instance, we have two ways of composing a sentence with a subject NP and a VP.

(14) $S \to NP\ VP$

a. $\lambda k.\,[\![NP]\!]\,(\lambda x.\,[\![VP]\!]\,(\lambda P.\,k\,(P\,x)))$

b. $\lambda k.\,[\![VP]\!]\,(\lambda P.\,[\![NP]\!]\,(\lambda x.\,k\,(P\,x)))$

(14a) corresponds to the left-to-right strategy and (14b) to the right-to-left strategy. Since the quantifier that is evaluated first contains another quantifier in its context, it takes the broader scope. Therefore, (14a) derives the surface scope reading, and (14b) derives the inverse scope reading.

Barker's definition can correctly account for sentences that contain two quantifiers. However, when a sentence has three quantifiers, some unfavorable readings are derived and some possible readings are not. Consider the sentence (15).

(15) Some teachers introduced most students to every company.

Following Barker, we can compose from left to right or from right to left both the sentence (first the subject NP, then VP, or vice versa) and the VP (first the object NP, then the prepositional phrase, or vice versa). Therefore, the following four readings are derivable:

$$\begin{aligned} &\mathit{some} > \mathit{most} > \mathit{every} \\ &\mathit{most} > \mathit{every} > \mathit{some} \\ &\mathit{some} > \mathit{every} > \mathit{most} \\ &\mathit{every} > \mathit{most} > \mathit{some} \end{aligned}$$

When a sentence contains three quantifiers Q_1, Q_2 and Q_3 in their linear order, Bekki and Asai (2010) claim that it rarely shows the reversed reading $Q_3 > Q_2 > Q_1$. In contrast, intermediate-inverse readings, such as $Q_2 > Q_1 > Q_3$ and $Q_3 > Q_1 > Q_2$, do exist. These two readings are not derivable in Barker's analysis.

Furthermore, in Barker's framework, all linguistic expressions, including the expressions that do not need to refer to their context, are written in CPS. We will show an alternative approach in which the representations of names and verbs remain in direct style.

5 Proposal

In this section, we show that our analysis of focus and inverse scope by `shift/reset` can successfully resolve the problematic phenomena discussed above. Our analysis is implemented in "OchaCaml", which is a variant of ML.

5.1 Implementation in OchaCaml

OchaCaml is a `shift/reset`-extension of Caml Light (Leroy 1997). Here is an example of a program written in OchaCaml:

```
# 1 + reset (fun () -> 2 * shift (fun k -> k (k 3))) ;;
- : int = 13
```

The `reset` operator receives a thunk (0-argument function) and evaluates the body in a delimited context. The `shift` operator captures the current continuation (in the case above the function $\lambda x.\,(2 * x)$) and binds it to `k`.

We write the semantic representations of sentences in OchaCaml and execute them. Our interest is in how the sentences that contain focus or inverse scope should be represented in order to be properly computed. For this purpose, we represent names like "John" and "Mary" as strings, and verbs (e.g., "love" and "introduce") and logical predicates (e.g., $\wedge$ and $\forall$) as functions that receive an appropriate number of arguments and return a string. Our implementation is enough to observe the behavior of `shift/reset`. We do not discuss

how sentences are mapped to their initial semantic representation. Note that the reduced representations (i.e., those which do not contain the control operators) are first-order formulas, so we are not concerned with how they are to be interpreted.

5.2 Focus with `shift/reset`

We follow the definition given in Bekki and Asai (2010) when the sentence does not contain nested focus. The sentence (3) discussed in Section 3 is represented in the following way:

(3) John only loves $[\text{Mary}]_{\text{F}_\text{o}}$.

```
# reset (fun () -> love
                   (shift (fun k -> forall x (k x <-> x = m))) j) ;;
- : string = "forall x (love (j, x) <-> x = m)"
```

The sentence (16) below has a reading which presupposes that there is someone distinct from Sue to whom Mary introduced only Bill. It has two focusing adverbs and two distinct foci. In this sentence, we expect the existential quantifier associated with "also" to take wider scope than the universal quantifier associated with "only". OchaCaml can reduce the representation in the desirable way because it evaluates the arguments right-to-left. Thus we obtain the representation in which the existential quantifier takes the wider scope. If the arguments are evaluated left-to-right, the universal quantifier will take the wider scope.

(16) Mary also only introduced $[\text{Bill}]_{\text{F}_\text{o}}$ to $[\text{Sue}]_{\text{F}_\text{a}}$.

```
# reset (fun () -> introduce
               (shift (fun k1 -> forall x (k1 x <-> x = b)))
               (shift fun k2 -> exists y (k2 y & not (y = s))) m) ;;
- : string = "exists y (forall x
                (introduce (m, x, y) <-> x = b) and not (y = s))"
```

In order to represent Krifka's nested focus, we defined layered control operators, namely `shift1/reset1` and `shift2/reset2` in OchaCaml. `shift1/reset1` behave in the same way as the standard `shift/reset`, and `shift2/reset2` are one level higher than them. `shift2` can capture the computation enclosed with `reset2` even if the surrounding context up to `reset2` contains `reset1` in its scope. The typical implementation of `shift2/reset2` is realized by writing a program in 2CPS (a program that is transformed twice into CPS). Indeed, we define them in direct style by making use of the static `shift/reset` of OchaCaml.

Using these operators, the representation of the sentence (9) can be properly reduced (for technical reasons, one has to enclose the entire representation with `run1 (fun () -> )`).

(9) Sue has also thought that John only loves $[[\text{Mary}]_{\text{F}_\text{a}}]_{\text{F}_\text{o}}$.

```
# run1 (fun () ->
   reset2 (fun () -> think
    reset1 (fun () -> love
     shift1 (fun k1 -> forall x (k1 x <->  x =
      shift2 (fun k2 -> exists y (k2 y & (not (y = m)))))) j) s)) ;;
- : string =
    "exists y (think (s,
      forall x (love (j, x) <-> x = y)) & not (y = m))"
```

We assigned `shift2/reset2` for the representations of "also" and its focus. Thus the `shift2` operator can "go across" the `reset1` associated with "only" and capture the continuation up to the predicate "think". As shown in (7), the adverb that takes the wider scope associates with the inner focus. This means that the adverb that appears first has to be represented with the `reset2` operator and its focus with the `shift2` operator, otherwise the `reset` operator associated with the second adverb would prevent the inner focus from capturing the proper context.

5.3 Inverse Scope with `shift/reset`

First, we define three operators of distinct types for each quantifier: quantifiers in subject position, object position, and prepositional phrase. In the definition below, `n` is some common noun, such as "man", `p` is some verb phrase, such as "run" or "loves Mary", `s` is the subject NP, and `o` is the object NP.

```
let every n p = "forall x (" ^ (n x) ^ " -> " ^ (p x) ^ ")" ;;
let every_acc n p s = "forall x (" ^ (n x) ^ " -> " ^ (p x s) ^ ")" ;;
let every_pp n p o s = "forall x (" ^ (n x) ^ " -> " ^ (p o x s) ^ ")" ;;
```

Then we define the inverse scope operator `inv` in terms of the `shift` operator:

```
let inv f = shift (fun k -> f k) ;;
```

The `inv` operator receives a quantified NP `f` that has the type of subject NP. The `shift` operator captures the context of `f`, and this context is passed to `f`. This results in a representation where the quantifier contained in `f` takes the broader scope. One has to enclose the whole representation with `reset` because the `inv` operator contains a `shift` operator.

Using the `inv` operator, we can write representations in direct style. The inverse scope reading of the sentence (12) is represented in the following way.

(12b) Some woman loves $[\text{every man}]_{\text{INV}}$.

```
# reset (fun () -> some woman (love (inv (every man)))) ;;
- : string = "forall x (man (x) ->
                  exists y (woman (y) & love (y, x)))"
```

The two intermediate-inverse scope readings of (15) are derivable by separately applying `inv` to "most students" and "every company".

(15a) Some teachers introduced [most students]$_{\mathrm{INV}}$ to every company.
(*most* > *some* > *every*)

```
# reset (fun () -> some teacher
   (every_pp company introduce (inv (most student)))) ;;
- : string = "most z (student (z), exists y (teacher (y) &
forall x (company (x) -> introduce (y, z, x))))"
```

(15b) Some teachers introduced most students to [every company]$_{\mathrm{INV}}$.
(*every* > *some* > *most*)

```
# reset (fun () -> some teacher (most_acc student
(fun s -> introduce s (inv (every company))))) ;;
- : string = "forall x (company (x) -> exists y (teacher (y) &
most z (student (z), introduce (y, z, x))))"
```

If one applies `inv` to both "most students" and "every company" simultaneously, the reversed reading *every* > *most* > *some* is derived. To avoid this, we set a restriction to allow only one `inv` operator in each sentence.

One might say that in Barker's analysis, surface/inverse scope readings are obtained in a more uniform way. We rather think that the requirement of the `shift` operator captures the nature of inverse scope. Inverse scope reading generally requires a process that is more complicated than surface scope reading. This fact is reflected in our analysis, because inverse scope readings are derived by using an additional operator.

6 Interaction between Focus and Inverse Scope

We considered sentences that contained both focus and more than two quantified NPs, and observed how the representations were reduced in OchaCaml. In this section, we will argue that in some cases, the quantifiers cannot take inverse scope. We will show two such examples.

First, consider the sentence (17).

(17) Some woman only introduced every man to [John]$_{\mathrm{F_o}}$.

Does (17) have the inverse scope reading? Let us see how the representation in which we apply the `inv` operator to "every man" is reduced. If we take the scope of "only" to be the whole sentence, the inverse scope reading of (17) will be represented as follows (here we abbreviate the representation of $[M]_{\mathrm{F_o}}$ to `f_o M`):

(18) Some woman only introduced [every man]$_{\mathrm{INV}}$ to [John]$_{\mathrm{F_o}}$.

```
# reset (fun () -> some woman
  (introduce (inv (every man)) (f_o j))) ;;
- : string = "forall x (forall y (man (y) -> exists z (woman (z) &
                          introduce (z, y, x))) <-> x = j)"
```

As mentioned above, OchaCaml evaluates the arguments right-to-left, so `f_o j` is evaluated first. Thus we obtain the representation in which the universal quantifier contained in the representation of the focus $[\text{John}]_{\text{F}_\text{o}}$ takes the widest scope. This representation means that the proposition "for every man, there exists some woman who introduced him to x" is true iff x is John. However, (17) seems not to have such a reading. Indeed, we think that in (17) "every" cannot take the inverse scope. To avoid this undesirable reading, we can limit the scope of "only" to the verb phrase. In the representation below, we enclose the verb phrase, rather than the entire sentence, with `reset`. Then, "every" cannot take scope over "some".

```
# some woman (reset (fun () ->
  introduce (inv (every man)) (f_o j))) ;;
- : string = "exists x (woman (x) & forall y (forall z (man (z) ->
                          introduce (x, z, y)) <-> y = j))"
```

This reading is "there is some woman who introduced every man to only John". Sentences in which the quantified NP intervenes between an adverb and its focus, as in (17), seem not to have the inverse scope reading. Probably, taking the inverse scope while combining the adverb and its focus is a process that is too complex for human beings.

Next, consider the case where a quantified NP is focused.

(19) Some teachers only introduced $[\text{most students}]_{\text{F}_\text{o}}$ to every company.

The intermediate-inverse scope reading *most* > *some* > *every* can be derived if one applies the `inv` operator to "most students".

(20) Some teachers only introduced $[[\text{most students}]_{\text{INV}}]_{\text{F}_\text{o}}$ to every company.

```
# reset (fun () -> some teacher (every_pp company
   introduce (f_o (inv (most student))))) ;;
- : string = "most x (student (x), forall y (exists z (teacher (z) &
       forall w (company (w) -> introduce (z, y, w))) <-> y = x))"
```

In `f_o (inv (most student))`, the `shift` operator in `f_o` is executed first. The context captured by the `shift` in `inv` is limited to the body of the first `shift`, and this context is passed to `most student`, resulting in a representation where "most" takes the widest scope. Thus, this is the reading "for most students, there exists some teacher who introduced only him or her to every company". However, (19) does not indicate such a situation. Again, we try to limit the scope of "only".

```
# some teacher (fun t -> reset (fun () -> every_pp company
   introduce (f_o (inv (most student))) t)) ;;
- : string = "exists x (teacher (x) & most y (student (y),
               forall z (forall w (company (w) ->
                introduce (x, z, w)) <-> z = y)))"
```

Here, the reading is "there exists some teacher who introduced only most students (not every student) to every company". We think that (19) does have such a reading.

To summarize, when a sentence contains a focus and more than one quantifier, the scope of "only" seems to be limited within the verb phrase, and thus the quantifier that has the `inv` operator applied cannot take scope over the subject NP.

7 Conclusion and Future Work

In this paper, we have proposed an analysis of focus and inverse scope by means of `shift/reset` and presented the implementation in OchaCaml. We showed that our approach can account for Krifka's nested focus and for sentences that contain three quantifiers. We also discussed the interaction between focus and inverse scope. We found that in some cases the reading derived by simply applying the rules that we have given for focus and inverse scope is not empirically preferred, and made a prediction that sentences which contain a focus hardly have inverse scope reading.

There are several issues that will be addressed in future work. Focus and inverse scope can interact in various ways. We have only discussed some typical cases in this paper, so we will examine more examples. Another issue is that in the current formulation, the definition of focus varies by adverbs, hence focus interpretation is not computed compositionally. We will develop a formulation in which focus is represented uniformly, and the meaning brought by an adverb is described in the representation for the adverb, so as to calculate the meaning compositionally.

Acknowledgements

I wish to thank three anonymous reviewers for their insightful comments, as well as Kenichi Asai, Daisuke Bekki, Koji Mineshima and Pascual Martínez-Goméz for helpful suggestions and discussions. This research was supported by JST, CREST.

References

1. C. Barker. Continuations and the nature of quantification. *Natural Language Semantics*, 10(3):211–242 (2002)
2. C. Barker. Continuations in Natural Language. *CW*, 4:1–11 (2004)
3. D. Bekki and K. Asai. Representing Covert Movements by Delimited Continuations. In K. Nakakoji, Y. Murakami, and E. McCready, editors, *New Frontiers in Artificial Intelligence (JSAI-isAI 2009 Workshops, Selected Papers from LENLS 6)*, volume LNAI 6284, pages 161–180. Springer (2010)
4. O. Danvy and A. Filinski. Abstracting Control. In *Proceedings of the 1990 ACM conference on LISP and functional programming*, pages 151–160. ACM (1990)
5. P. de Groote. Type raising, continuations, and classical logic. In R. van Rooij and M. Stokhof, editors, *Proceedings of the thirteenth Amsterdam Colloquium*, pages 97–101 (2001)
6. M. Krifka. A Compositional Semantics for Multiple Focus Constructions. In S. Moore and A. Z. Wyner, editors, *Proceedings of SALT*, volume 1, pages 127–158 (1991)
7. X. Leloy. The Caml Light system release 0.74. *URL: http://caml. inria. fr* (1997)
8. M. Masuko and K. Asai. Caml Light + shift/reset = Caml Shift. *Theory and Practice of Delimited Continuations (TPDC 2011)*, pages 33–46 (2011)
9. R. May. *The Grammar of Quantification.* PhD thesis, Massachusetts Institute of Technology (1977)
10. G. D. Plotkin. Call-by-Name, Call-by-Value and Lambda Calculus. *Theoretical Computer Science*, 1(2):125–159 (1975)
11. M. Rooth. A Theory of Focus Interpretation. *Natural Language Semantics*, 1(1):75–116 (1992)
12. M. Rooth. *Focus*, pages 271–298. Blackwell (1996)
13. M. Wagner. NPI-Licensing and Focus Movement. In E. Georgala and J. Howell, editors, *Proceedings of SALT*, volume 15, pages 276–293 (2006)

Toward an Ontology-Based Chatbot Endowed with Natural Language Processing and Generation

Amine Hallili

Univ. Nice Sophia Antipolis, CNRS, I3S, UMR 7271, 06900 Sophia Antipolis, France
hallili@i3s.unice.fr

Abstract With the last evolution of the web, several new means of communication have showed up. In the commercial domain, chatbot technologies are now considered as essential for providing a wide range of services (e.g. search, FAQ, assistance) to the end-user, and to make a client a faithful customer. In this paper, we propose an on-going work on the definition and implementation of SynchroBot, an ontology-based chatbot that relies on Semantic Web and NLP models and technologies to support user-machine dialogical interaction in the e-commerce domain.

Keywords: Chatbot, Artificial Intelligence, Natural Language Processing, Natural Language Generation, Semantic Web

1 Introduction

During the last decades, our way of consuming information has totally changed with the emergence of new means of communication (e.g. forums, FAQ, social networks, semantic search engines, mobile applications, and text to speech systems) which provide us with different possibilities of handling and dealing with information on the web. At the same time, researchers in Natural Language Processing (NLP) and Semantic Web domains have proposed new approaches to model and implement more and more complex systems capable of interpreting natural language, of reasoning, and of assisting end-users (e.g. Chatbots [1], Expert Systems [10] , multi agent systems [15], and Question Answering systems [9]). Besides covering both open and close domains (e.g. social, commercial, scientific), such systems aim to be autonomous, self-learning and they can replace humans in performing several tasks. My PhD research proposal, whose preliminary works I present in this paper, focuses on chatbot systems, which we classify in two different categories: Question Answering Systems (QA) and Dialog Systems (DS). On the one hand, Question Answering systems aim at finding answers to factual queries in either a Knowledge Base (KB) or raw text and to return them to the user. The answer can be just a textual string (e.g. [4]) or it can be enriched by other meta-information or well-formed sentences, obtained by applying Natural Language Generation (NLG) techniques (e.g. [2,5]). In spite of their efficiency in retrieving the information, such systems lack the capability of handling the links between sequential questions as in a conversation. On

the other hand, Dialog Systems aim at keeping in memory the links between consecutive questions in order to ensure a logical conversation mode with the user (e.g. [13,7]). Nevertheless, most of these systems do not rely on robust and flexible KBs allowing them to extract information from multiple sources and to reason over the data. The goal of our work is to combine the strenghts of the two categories of systems discussed above, and to propose a dialog system that relies on i) a rich KB for data extraction and reasoning, ii) NLP tools to interpret user's question, and iii) NLG techniques to generate well-formed sentences. The system will ensure the following type of conversation:

<User> *Give me the price of a Nexus 5!*
<System> *the price of Nexus 5 is 400$*
<User> *and who sells it?*
<System> *several sellers were found. The main one is Google! Do you want to see other sellers?*
<User> *No, show me the white version, sold by Google and located in France!*
<System> *here are the images of Nexus 5 white version, sold by Google and located in France...*

The remainder of this paper is organized as follow: Section 2 presents our preliminary approach and implementation. In Section 3 we describe our ongoing works along with our perspectives for future works.

2 SynchroBot: A Preliminary Approach

The approach we propose relies on the Semantic Web[1] paradigm, which covers structuring, linking, sharing and reusing data through applications, enterprises and communities. For that, it provides a number of information modeling frameworks e.g. Resource Description Framework (RDF) and RDF Schema (RDFS). The preliminary approach we propose here toward an ontology-based chatbot covers three aspects, namely i) Knowledge based System ii) Question Interpretation iii) Natural Language Generation. Currently we focus on modeling and implementing an efficient and robust QA system that will be the corner stone for our future Dialog System.

2.1 A Knowledge Based System

Our approach relies on the use of exiting tools, resources and information (e.g. FAQ, API, system logs) in order to create a KB in RDF. For example, the following sentence 'Google sells Nexus 5' can be represented by the following RDF triple `<sbr:Google, sbo:sells, sbr:Nexus_5>`. We have created an ontology that describes the classes (e.g. Product, Category, Seller, etc.) and properties

[1] www.w3.org/2001/sw/

(e.g. sells, price, locatedIn, etc.) of the KB in the e-commerce domain (the focus of Synchrobot). For instance `sbr:Google` is of type `sbo:Seller` and is the subject of a `sbo:sells` property.Every property is annotated in both French and English, by a number of labels (e.g. `sbo:sells` will have 'sell', 'trade', 'vend', 'commercialize', 'market', etc. as labels), which will be used to match the terms in the question, in order to identify the queried property. Some algebraic properties are defined in the ontology enabling inferences; for instance, `sbo:sells` is defined as the inverse property of `sbo:soldBy` which enable to infer that Nexus 5 is sold by Google (`<sbr:Nexus_5`, `sbo:soldBy`, `sbr:Google>`). The current version of the KB is composed of 500000 product descriptions that we retrieved by using eBay APIs to transform eBay data to RDF.

2.2 Question Interpretation

As regards the natural language question interpretation, our approach focuses on textual information as input and relies on the work described in [3] which requires identifying three aspects: i) the Expected Answer Type (EAT), which is the type of the resource that we are looking for, ii) the property, representing the relation linking the entity on which the question is asked to its answer, and iii) the Named Entity (NE) representing the subject of the given question. In this example 'Who sells Nexus 5?', the EAT is `sbo:Seller`, the property is `sbo:sells` and the NE is `sbr:Nexus_5`.

Named Entity Recognition: To identify the NE, we aim at using natural language processing techniques (e.g. Named Entity detection and linking) to retrieve all possible NEs from the KB by matching the user'question to KB property values (e.g. name, description, etc.). Then, relevant NEs will be used in querying the KB according to scores which we assign by using the folowing strategy. First, all KB properties used for the search are ordered depending on their relevance. For instance, matching the user's question to the `sbo:hasLegalName` property will be more accurate than matching it to the `sbo:hasDescription` property. Based on that, a relevance coefficient is assigned to each property and used in both the retrieval and scoring of relevant NEs. Second, a score is assigned to the accuracy of the matching of the user's question with the resources found in the KB. The more matches there are, the higher the score will be (an exact match will result of a maximal score and the found resource will be directly used). Finally, we focus on the number of retrieved resources to determine the precision of our result. This means that the fewer resources we find, the higher the precision of the search.

Property Detection: We detect the property involved in a question by matching its labels with the words in the user's question [6] and following a scoring strategy we pick up the relevant property.

We are also able to recognize questions with two relations as shown below in figure 1. For instance, the following question 'Give me the address of the Nexus

5's sellers!' contains two properties, `sbo:address` and `sbo:sells`, meaning that the question can be divided in two sub-questions: 'Give me the Nexus 5's sellers!' and 'Give me their addresses!'. This can be done by using the property domains and ranges stated in our ontology. Concretely we aim at constructing a relational graph representing the user's question that contains the identified properties along with the found resources (e.g. Named Entities), while comparing both the NE type and the identified property domain. In the given example, the property `sbo:address`, with domain `sbo:Seller`, will have the best score along with the NE `sbr:Nexus_5` which is of type `sbo:Product` which differs from `sbo:Seller`. This leads to the creation of a relational graph with two property nodes, namely `sbo:address` and `sbo:soldBy`.

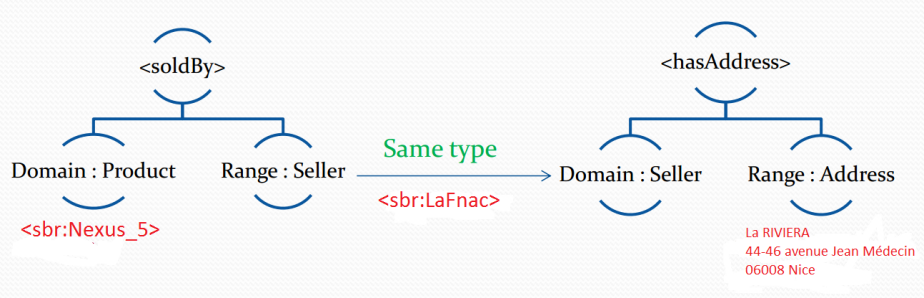


Figure 1: Relational graph for 2 relations

Expected Answer Type (EAT) Detection: After detecting relevant properties, we aim at using their respective `domain`. For instance, as shown in figure 1, the following property `domains` (`sbo:Product`, `sbo:Seller`, `sbo:Address`) will be used along with the score assigned to their respective property. This allows us to sort all the detected EATs before adding them in the generated query that the system will use to retrieve results from the KB enabling to construct an answer to the user question.

2.3 Natural Language Generation

In order to answer questions with a generated sentence, we propose that each property in our ontology will be mapped with a list of generic response patterns. Our challenge is to be able to replace dynamically some particular parts of the pattern to return well-formed answers. For instance, we take the example 'Give me the price of a Nexus 5!' and considering that the identified property `sbo:price` matches the pattern *[The price of {Product} is {Value}]*, so after replacing the {Product} and {Value} parts we can answer that 'The price of Nexus 5 is 400$'. We also use the `sbo:mediaType` property to show more interesting information to the user after giving the well-formed answer (e.g. image,

video, map, etc.). For instance, when a user asks 'Show me the white version of Nexus 5!', the system infers that the user is more interested in viewing images rather than just textual information, and as a result, images will be displayed in this case.

3 Ongoing and Future Work

This paper sketches out the PhD research project I have recently started, and describes the preliminary steps representing the bases of a number of directions that we consider for future work:

As ongoing and short term works we are planning to improve the NE recognition by using well-known and efficient algorithms (e.g. KNN, Similarity, N-Gram and TF-IDF scoring) in order to gain more precision, to spot complex and ambiguous resources, and to be able to diversify given answers. We have also begun to reuse other famous ontologies that exist in the literature and cover the commercial domain. we have started to use the schema.org [11] ontology but due to its parial coverage, we have decided to use more specific ontology (GoodRelations [8] ontology) which fully covers the commercial domain.
Furthermore, we consider answering n-Relation questions by the construction of a Relational Graph representing the question's NEs and properties. For that, we will study the possibility to generalize the 2-relation question answering approach explained in the previous section, to n-Relation questions. Moreover, we intend to integrate the relational pattern matching module of QAKiS system [4] that exploits Wikipedia pages to extract lexicalisations of ontological relations. More specifically, we will use website APIs, web services [14] and product pages to automatically extract and create generic property and response patterns. This will ensure more precision in detecting properties expressed in the user's question and it will allow to answer questions in different ways.

As middle term improvements, we intend to focus our work on the dialog mode part that we will integrate on top of our proposed approach, so that we can propose an approach that ensures our targeted scenario. For that, we will investigate the communicative behavior approaches (e.g. pause, resume, and to switch between interactive tasks [13]), dialog management systems (e.g. [7]) and in particular, the ontology-based dialog systems (e.g. [12] which correspond perfectly to the kind of system that we want to implement.

References

1. Allen, J.F., Byron, D.K., Dzikovska, M., Ferguson, G., Galescu, L., Stent, A.: Toward conversational human-computer interaction. AI Magazine 22(4), 27–38 (2001)
2. Augello, A., Pilato, G., Vassallo, G., Gaglio, S.: A semantic layer on semi-structured data sources for intuitive chatbots. In: CISIS. pp. 760–765 (2009)
3. Cabrio, E., Cojan, J., Aprosio, A.P., Magnini, B., Lavelli, A., Gandon, F.: Qakis: an open domain qa system based on relational patterns. In: International Semantic Web Conference (Posters & Demos) (2012)
4. Cabrio, E., Cojan, J., Palmero Aprosio, A., Gandon, F.: Natural language interaction with the web of data by mining its textual side. Intelligenza Artificiale 6(2), 121–133 (2012)
5. Chai, J., Horvath, V., Nicolov, N., Stys, M., Kambhatla, A., Zadrozny, W., Melville, P.: Natural language assistant: A dialog system for online product recommendation. AI Magazine 23, 63–75 (2002)
6. Damljanovic, D., Agatonovic, M., Cunningham, H.: Natural language interfaces to ontologies: Combining syntactic analysis and ontology-based lookup through the user interaction. In: ESWC (1). pp. 106–120 (2010)
7. Heinroth, T., Denich, D.: Spoken interaction within the computed world: Evaluation of a multitasking adaptive spoken dialogue system. In: COMPSAC. pp. 134–143 (2011)
8. Hepp, M.: Goodrelations: An ontology for describing products and services offers on the web. In: EKAW. pp. 329–346 (2008)
9. Hirschman, L., Gaizauskas, R.J.: Natural language question answering: the view from here. Natural Language Engineering 7(4), 275–300 (2001)
10. Liao, S.H.: Expert system methodologies and applications - a decade review from 1995 to 2004. Expert Syst. Appl. 28(1), 93–103 (2005)
11. Mika, P., Potter, T.: Metadata statistics for a large web corpus. In: LDOW (2012)
12. Milward, D., et al.: Ontology-based dialogue systems (2003)
13. Pakucs, B.: Towards dynamic multi-domain dialogue processing. In: INTERSPEECH (2003)
14. Sonntag, D., Engel, R., Herzog, G., Pfalzgraf, A., Pfleger, N., Romanelli, M., Reithinger, N.: Smartweb handheld - multimodal interaction with ontological knowledge bases and semantic web services. In: Artifical Intelligence for Human Computing. pp. 272–295 (2007)
15. Wooldridge, M.J.: An Introduction to MultiAgent Systems (2. ed.). Wiley (2009)

A Two-Step Scoring Model for Computational Phylolinguistics

Nancy Retzlaff

Bioinformatics Group, Department of Computer Science
University of Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany.

Abstract In computational phylolinguistics, methods usually used for sequence alignment and phylogenetic tree reconstruction are transferred to historical linguistics. In this paper, a two-step scoring model for performing linguistic word alignments is proposed. Based on an initial unigram-based model, a more complex bigram-aware model is developed and trained. In addition, a phylolinguistic tree of language distances is computed which conforms well to other, established models of language relations.

1 Introduction

With more than 7 000 different languages [1], language diversification has turned into one of the most interesting facts about rebuilding part of the history of humankind. Phylolinguistics, a computational approach where phylogenetic approaches are used to study historical linguistics, brings a great opportunity to address the old question of how languages are related to each other. It is already well known that Indo-European languages share the same root and that descendents of the Proto-Indo-European language are the consequence of variation on that root, as first postulated by Gottfried Wihelm Leibniz (1646 – 1716) [2]. However, it was not until 1822, when Jacob Grimm (1785 – 1863) used a comparative approach, that this discovery was scientifically accepted [3]. Nonetheless, language diversity still opens the opportunity to explore questions related to the emergence of new languages, but also to the extinction or disappearance of others [4], to classification [5], and to language universals [6]. The comparative method used by Jacob Grimm bears a similarity to those for the alignment of biological sequences such as DNA, RNA, and amino acids. Therefore, borrowing and applying phylogenetic methods in a historical linguistics context (as done in [7,8,9]) opens up new ways of automatically dealing with a huge amount of data. A general overview of common ground between historical linguistics and biology can be found in [10].

In this approach the Needleman-Wunsch algorithm [11] is used. As a global alignment algorithm it considers the whole sequences, usually DNA or amino acids in biology. Substituting these biological components by words now allows to compare words, as already done in a distance-based approach in [12], which measured the discrepancy of words. This paper will focus on the similarity of

words which is motivated by the calculation of log-odd scores, further introduced in Sec. 2 (see below). In fact, the famous BLOSUM [13] and PAM [14] scoring matrices used in computational biology have been obtained in this manner. These scores offer a likelihood of exchanges in a sequence what makes them a possible measure of similarity between characters and, hence, words.

The idea here has an evolutionary perspective. In molecular biology, sequence evolution is usually approximated by a stationary, reversible Markov process [15, Cha. 14]. Thus, scoring models are symmetric and character identity is the most suitable predictor for homology, i.e., matches always score better than mismatches. From a linguistic perspective, however, these assumptions are still not satisfied. This can be due to the fact that one can find more regular phenomena occurring, for example, in Indo-European languages. Some regular phenomena one can find in this language family were categorized under the term Grimm's law [3,16], which describes sound shifts regularly occurring when comparing Indo-European languages.

Sound shifts make it almost impossible to use a symmetric model of language change, since a sound shift is unidirectional. In terms of the word alignment model considered in this paper, this marks the most important distinction to be made when borrowing alignment methods used in computational biology.

Alignment algorithms for biological sequences usually assume that alignment columns are independent. Such columns can be seen in the example alignment (Fig. 1). Again, this is not a particularly good approximation for linguistic data since the syllable structure of words may have a strong influence on sound changes and may make them context dependent [17,18]. Working with linguistic data raises the question of how to separate words in syllables without actually knowing the boundaries of those. An approach to automatic syllabification [19] tried to solve the problem by considering different n-grams. The results showed that the accuracy increases logarithmically with a larger n. Returning to the scoring model for words, it can easily seen that there is a trade-off between the problem of the best syllable prediction and the sparse data problem. Hence, a further innovation in this approach is stepping away from the usual character or unigram based models to a bigram scoring model, a model that considers two adjacent characters. (For example the word 'bigram' yields the bigram set {'^ b', 'bi', 'ig', 'gr', 'ra', 'am', 'm$'}. Here, the symbols ^ and $ are used to mark the beginning and end of a word as shown in Fig. 1.) For linguistic data, a first step was taken without considering the order of those bigrams [20]. In biology, one can also find bigram-based methods in order to create a context via considering adjacency [21].

Measuring similarities now enables addressing the problem of cognate detection or classification of words, and thus words with a common etymological origin. Furthermore, this information can also be used for building a phylolinguistic tree.

Recapitulatory, the following ideas are developed in this paper:
(I) A simple unigram-based scoring model is given, that is used to perform an *initial alignment* between words with the same meaning for all pairs of languages.

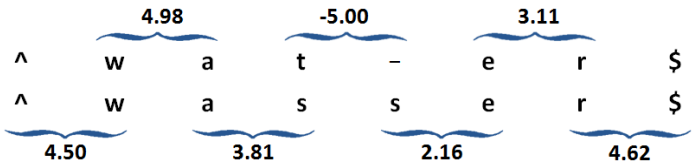


Figure 1: Example of a bigram alignment for the words 'water' (engl.) and 'wasser' (germ.) with special annotation of the bigram scores marked by curly brackets. Each bigram pair of both words is scored separately. In the conventional Needleman-Wunsch algorithm each unigram pair would be scored individually. The sum of all pair scores yields the score 18.18 for the whole alignment.

This is done in lieu of cognate alignment.
(II) Using the initial alignment, a more complex scoring model based on *bigram frequencies* is derived.
(IV) All words (i.e. assumed cognates and non-cognates) are realigned with a more complex bigram-aware grammar.
(IV) Finally, phylogenetic trees are constructed based on the bigram alignments.

2 Methods

This approach is based on word lists along with their meaning. Here, the word lists from the *Intercontinental Dictionary Series* [22] for (*i*) Breton, (*ii*) Danish, (*iii*) Dutch, (*iv*) English, (*v*) French, (*vi*) German Standard, (*vii*) Greek, (*viii*) Italian, (*ix*) Latin, (*x*) Spanish, and (*xi*) Swedish are used. As a version of Buck's list [23], the *IDS* lists are organized in 23 chapters with 1 310 semantic concepts. Semantic concepts can have one or more lexical items and, in case this concept does not exist in a language, it is left blank. The lexical items in this list may be written orthographically or phonetically and can even be (partial) orthographic transcriptions. For example, German 'ß' is transcribed as 'ss'. In some cases there is also information on stress. Hence, it is necessary to transform all words into a common, most compact, notation.

The first step taken is to remove diacritics and word stresses to reduce the amount of possible bigrams but still preserving as much information as possible. Therefore the Spanish ú becomes u (as in "e s t ú p i d o"). Further, German *umlaute* are represented as a vowel plus a diacritic. Those are normalized to a single *umlaut* character (e.g. ä encoded as Unicode symbols U+0061 (a) with U+0308 (¨) are replaced by U+00E4 (ä) so the German word for girl, "m a ¨ d c h e n", becomes "m ä d c h e n").
Unigram-based alignments. Determination of the cognacy relation in a computational approach is based on the similarity of words and therefore, on the alignments of words. To be able to align the words, a score model that contains information about the best matching bigram pairs has to be defined. Here, the

score model is defined on co-occurrences of bigrams. To reduce the majority of noise, only words with the same meaning are considered.

In [24], the log-odd scores were computed directly with the given data. As not all words with the same semantic concept are actual cognates, wrong signals could not be avoided. In addition, the results showed that the prothesis of 'e' in Spanish could not be considered during the alignment step. So an insertion of 'e' at the beginning of a sequence was hard to classify correctly. This happened due to the fact that the scoring model did not contain any information for this feature.

To avoid this type of problems, an initial alignment step is taken. It aligns unigrams regarding their nature as consonant, liquid, or vowel. In addition, equivalence classes for phonetically very similar characters are introduced.

This initial alignment system gives highest scores to equal characters (or those within the same equivalence class, i.e. 'ä' and 'a' denote the same character in the initial alignment). Characters within the same set of vowel, consonant, or liquid are still scored favorably. Other matches are penalized or outright disallowed, such as aligning a vowel to a consonant. Matching scores are denoted by 'x/y' (Table 1), where 'x' is the sound class.

Table 1: Chosen unigram scores for pre-alignment. Here $-\infty$ was used to avoid the alignment of consonants with vowels. Furthermore, an exchange between vowels is considered more likely then an exchange between consonants. This leads to the fact that the information content of aligning two vowels is smaller than compared to aligning two consonants.

	consonant	vowel	liquid
consonant	4 / 0	$-\infty$	-1
vowel	$-\infty$	2 / 0	-1
liquid	-1	-1	3 / -1

For the program, the ADPfusion framework [25] is used. It only needs (*i*) a grammar that represents all possible editing operations in order to produce aligned output sequences from the input sequences and (*ii*) a scoring model which contains scores for the evaluation process of the alignments.

Bigram frequency calculation. In order to compute the scores, the steps taken here are quite similar to the ones for the BLOSUM matrices [13]. With the pre-aligned words, it is possible to count co-occurrences of corresponding bigrams. It is assumed that words with the same meaning are very likely to be also cognates and to contain the regular sound correspondences. Thus, only alignments of words with the same meaning are used for calculation of the bigram scores. Given Equ. (1) the relative frequency of each bigram α in language A, written $\rho_A(\alpha)$ can be easily computed using the observed occurrences $\mathrm{occ}_A(\alpha)$. At the same time the occurrences of pairs of bigrams $\mathrm{occ}_{A,B}(\alpha, \beta)$ for bigrams α and β in languages A and B are counted as well. Which bigram pairs are

considered is determined by the pre-aligning step as only aligned bigram pairs are counted. Their relative frequencies $\rho_{A,B}(\alpha, \beta)$ are computed in Equ. (2) analogously to Equ. (1).

$$\rho_A(\alpha) = \frac{\mathrm{occ}_A(\alpha)}{\sum_\beta \mathrm{occ}_A(\beta))} \tag{1}$$

$$\rho_{A,B}(\alpha, \beta) = \frac{\mathrm{occ}_{A,B}(\alpha, \beta)}{\sum_{\gamma,\delta} \mathrm{occ}_{A,B}(\gamma, \delta)} \tag{2}$$

Taking those relative frequencies, it is easier now to compute log-odd scores $\sigma_{A,B}(\alpha, \beta)$ for each observed bigram pair (α, β) in every pair of languages A and B as shown in Equ. (3):

$$\sigma_{A,B}(\alpha, \beta) = \log \frac{\rho_{A,B}(\alpha, \beta)}{\rho_A(\alpha)\rho_B(\beta)} \tag{3}$$

Unfortunately, there are some bigram pairs which occur quite rarely and are probably alignment artifacts. They do, however, get a high score in this model as these might be non-cognates with the same meaning. To avoid this kind of noise only bigram pairs which are counted at least three times are taken into account. Those pairs that do not occur in the scoring model get a default score during the alignment that is less than the calculated scores.

Realignment. Given the bigram pair scores, it is now possible to realign all pairs of words. The bigram-aware grammar is, in addition, extended to handle affine gap costs and simple prothesis using a formal grammar framework [25,26,27]. In a previous work [24], prothesis happened to be a problem due to the calculation of the scores. These are simply computed via counting bigram pairs of two words with the same meaning. A common prothesis in Spanish is 'e' in words with leading 'st' as in 'estrella' (engl. *star*). Here it appears that words beginning with '^ e s t' in Spanish and words with leading '^ s t' sequences are scored so well that misleadingly '^ e' is going to be aligned with '^ s'. An example for the solution of this problem is shown in Table 2.

Phylolinguistic tree construction. In order to build a phylolinguistic tree, it is necessary to calculate similarity scores between the chosen languages. This can easily be done by calculating the average normalized alignment score of alignments with the same meaning. In contrast to biological substitution matrices [13,14], the matrices calculated here do not yield a universal scoring system, but rather score individual language pairs. However, all scores are calculated within the same log-odds framework and are, thus, still comparable.

To create the actual phylolinguistic tree, the similarities are first transformed into distances using the R package 'proxy' [28]. Then, using the 'ape' [29] package the Neighbour-joining [30] tree is created. Finally, Dendroscope [31] is used for visualizing the tree.

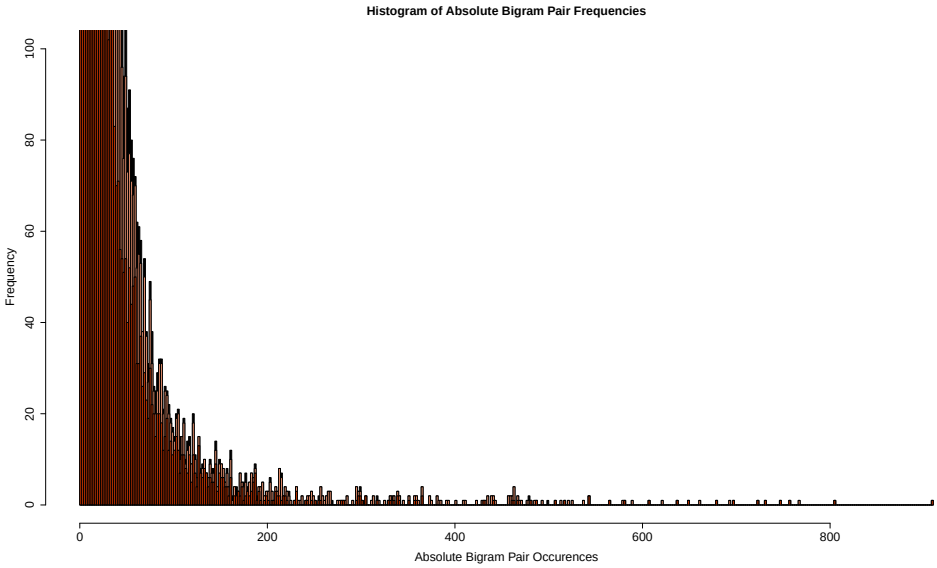


Figure 2: This histogram of absolute bigram pair frequencies shows the frequencies of different kinds of bigram pairs dependent on how often these occur. Identical bigrams got marked with orange (for example '‿ w' and '‿ w' as first bigram pair in Fig. 1). The colour salmon represents all bigram pairs that have one mismatch (for example 'a t' and 'a s' in Fig. 1) and all mismatching bigram pairs are shown in black (for example 't -' and 's s' in Fig. 1). To be able to see bigram pair occurrences that do not appear often, the histogram had to be cut at frequencies over 100. The highest frequency arises for bigram pairs that only occured once with a value of 307 381.

3 Results

Qualitative Analysis. Before starting the evalutation already mentioned in Sec. 1, a qualitative analysis of the computed absolute bigram pair scores is done. These counts are the occurrences necessary to calculate the relative frequencies of Equ. (1). In Fig. 2 a histogram of bigram pair frequencies is shown. On closer inspection one can see that the amount of mismatching bigrams in a bigram pair is almost not represented by bigram pairs that occur more often than 200 times. This fact supports the assumption made for the pre-alignment step that only bigram pairs which show similarity contain the relevant information.

Cognate detection. Cognates are words that share the same root in their formation. Whether a word has cognates is visually evaluated by simply considering the ten best alignment scores. An automated approach is introduced in Sec. 4.

In Table 2 the multiple alignments for the examples 'father' and 'star' are shown. Since a tool for this job has yet to be developed, the pairwise alignments played the pivotal role in constructing the multiple sequence alignment in Table 2. All words originated in the Proto-Indo-European language.

Table 2: Multiple alignment for eleven languages. The alignments are for concepts *father* and *star*. (Latin words for *father* and *star* are not contained in *Intercontinental Dictionary Series*). As one can also see here, the aforementioned problem of prothesis 'e' in Spanish is not a problem anymore. This is due to the fact that the scoring model now correctly contains bigram pairs for a deletion of 'e' in other languages.

Language	Alignment *father*	Alignment *star*
Spanish	p a d - r e -	e s t r - - - e l l a
Italian	p a d - r e -	- s t - - - - e l l a
Latin	p a t e r - -	- s t - - - - e l l a
French	p e r e - - -	e - t - - - - o i l e
Greek	p a t e r a s	a s t r - - - o - - -
German Standard	v a t e r - -	- s t - - - - e r n -
Dutch	v a d e r - -	- s t - - - - e r - -
Danish	f a d e r - -	- s t j - - - e r n e
Swedish	f a d e r - -	- s t j - - - ϵ r n a
English	f a θ e r - -	- s t - - - - a r - -
Breton	t a d - - - -	- s t e r e d e n n -

Sound shifts. The score for an alignment is simply the sum of all log-odd scores for each bigram pair. This means that sound shifts can be detected by aligning cognates as done in Table 2. Regular sound shifts following Grimm's law can be observed as they most likely occur in the same alignment column. For example $p \rightarrow f/v$, where v is the orthographic form vor f in relevant languages, or $t \rightarrow d \rightarrow \theta$.

Phlylolinguistic tree. The visualized phylolinguistic tree in Fig. 3 separates the set of languages into distinct subclasses. The Italic subfamily, which is represented here by French, Latin, Italian, and Spanish, can be found on the bottom (blue). On its left there is Greek (in light blue) which stands for its own language family. Breton belongs to the Celtic subgroup and was placed to the right (brown). Its root to the Germanic subfamily (orange), represented by English, German Standard, Dutch, Danish, and Swedish, is more recent than the roots to the other subgroups (compare to [8]). Returning to the Italic subfamily, it is known that Latin is a root for Italian and Spanish. This model does not consider Latin as an explicit ancestral language, but rather considers all languages independent of their historical context and estimates their neighbourhood relations. Thus, while Latin should be a root language for both Italian and Spanish, it simply occurs as a related language.

Bigram-based word alignments with an initial alignment step based on a very simple scoring model show a number of advantages in practise. Even though this model could be considered preliminary, it is already possible to (*i*) correctly determine cognates in other languages, (*ii*) identify regular sound shifts, and (*iii*) build phylolinguistic trees of languages that conform well to reality.

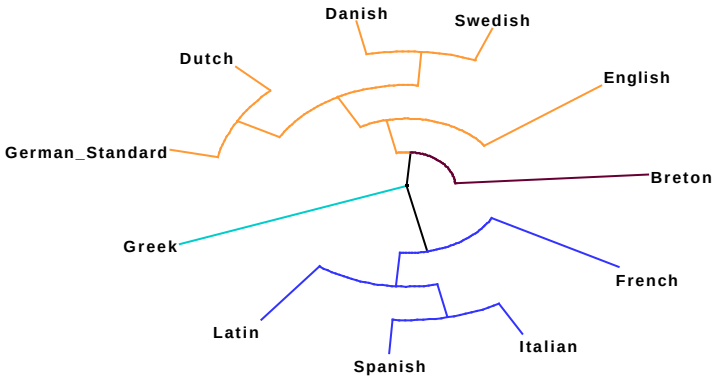


Figure 3: Phlylolinguistic tree for all chosen languages computed by applying the Neighbor-joining algorithm [30] and visualized by means of [31].

4 Conclusion and Further Work

Combining simple pre-alignments with a bigram-based scoring system yields encouraging results. For example, special linguistic phenomena like protheses can now be correctly aligned (compared to [24]). For the phylolinguistic tree (Fig. 3), a good benchmark is the tree of Atkinson and Gray [8]. Although they considered a larger number of languages in their approach, one can see obvious similarities between both trees.

The results also point into directions for further research. For instances, how to deal with both overfitting and missing data in the form of missing bigrams, which makes it necessary to improve the scoring model. A number of smoothing models have been proposed [32] and will be tested for this approach.

The underlying grammatical framework [26,27] will make it possible to consider more complicated grammatical operations such as metatheses without the need to encode these explicitly in the scoring model.

In order to detect cognates a cut-off has to be chosen. This can be done by considering the distribution of the alignment scores. Another possibility could be to calculate the false discovery rate and minimize it.

Based on this, a further advanced method could be automated cognate classification. This task concerns the grouping of words together in sets that should share the same root. The sets then can be built due to the identification of words that have the best alignment scores to each other.

Acknowledgements

Special thanks to Christian Höner zu Siederdissen, Lydia Steiner, and Peter F. Stadler who supported me in the process of writing this paper and even more for

making this possible. I also would like to thank the three anonymous reviewers for their kind and constructive criticism. Also thanks to my beloved sister for having a look. Last but not least, thanks to Alvaro Perdomo (Varo) for helping with issues of consistency and supporting me as a friend.

References

1. Lewis, M.P., Simons, G.F., Fennig, C.D.e.: Ethnologue: Languages of the World, Seventeenth edition. http://www.ethnologue.com (2013)
2. Arens, H.: Sprachwissenschaft. Der Gang ihrer Entwicklung von der Antike bis zur Gegenwart. Orbis Academicus. Freiburg/München (1969)
3. Gamkrelidze, T.V., Ivanov, V.: The early history of Indo-European languages. Scientific American **262**(3) (1990) 110–116
4. François, A.: Trees, Waves and Linkages: Models of Language Diversification. The Routledge Handbook of Historical Linguistics (2013)
5. Robinson, L.C., Holton, G.: Internal classification of the Alor-Pantar language family using computational methods applied to the lexicon. Language Dynamics & Change **2** (2013) 123–149
6. Levinson, S.C., Gray, R.D.: Tools from evolutionary biology shed new light on the diversification of languages. Trends in cognitive sciences **16**(3) (2012) 167–173
7. Jäger, G.: Evaluating distance-based phylogenetic algorithms for automated language classification. Technical report, Univ. Tübingen (2014) http://www.sfs.uni-tuebingen.de/~gjaeger/publications/njFastme.pdf.
8. Gray, R.D., Atkinson, Q.D.: Language-tree divergence times support the Anatolian theory of Indo-European origin. Nature **426**(6965) (2003) 435–439
9. Dunn, M., Terrill, A., Reesink, G., Foley, R.A., Levinson, S.C.: Structural phylogenetics and the reconstruction of ancient language history. Science **309**(5743) (2005) 2072–2075
10. Atkinson, Q.D., Gray, R.D.: Curious parallels and curious connections–phylogenetic thinking in biology and historical linguistics. Systematic biology **54**(4) (2005) 513–526
11. Needleman, S.B., Wunsch, C.D.: A general method applicable to the search for similarities in the amino acid sequence of two proteins. Journal of molecular biology **48**(3) (1970) 443–453
12. Jäger, G.: Phylogenetic inference from word lists using weighted alignment with empirically determined weights. Language Dynamics and Change (2013)
13. Henikoff, S., Henikoff, J.G.: Amino acid substitution matrices from protein blocks. Proceedings of the National Academy of Sciences **89**(22) (1992) 10915–10919
14. Dayhoff, M.O., Schwartz, R.M., Orcutt, B.C.: A model of evolutionary change in proteins. In Dayhoff, M.O., ed.: Atlas of Protein Sequences and Structure. Volume 5. Natl. Biomed. Res. Found., Silver Springs, MD (1978) 345–352
15. Ewens, W.J., Grant, G.R.: Statistical Methods in Bioinformatics: An Introduction. Volume 746867830. Springer (2005)
16. Campbell, L.: Historical linguistics: An introduction. MIT press (1998)
17. Vennemann, T.: Preference Laws for Syllable Structure: And the Explanation of Sound Change with Special Reference to German, Germanic, Italian, and Latin. Walter de Gruyter (1987)
18. Murray, R.W., Vennemann, T.: Sound change and syllable structure in germanic phonology. Language (1983) 514–528
19. Bartlett, S., Kondrak, G., Cherry, C.: Automatic Syllabification with Structured SVMs for Letter-to-Phoneme Conversion. In: ACL. (2008) 568–576
20. Rama, T., Borin, L.: Estimating language relationships from a parallel corpus. A study of the Europarl corpus. In: NEALT Proceedings Series (NODALIDA 2011 Conference Proceedings). Volume 11. (2011) 161–167

21. Bussotti, G., Raineri, E., Erb, I., Zytnicki, M., Wilm, A., Beaudoing, E., Bucher, P., Notredame, C.: BlastR fast and accurate database searches for non-coding RNAs. Nucleic acids research **39**(16) (2011) 6886–6895
22. Key, M.R., Comrie, B.: Intercontinental dictionary series. `http://lingweb.eva.mpg.de/ids/` (2007)
23. Buck, C.: D. 1949. a dictionary of selected synonyms in the principal indo-european languages (1970)
24. Retzlaff, N.: Bigramm-Alignierung und ihre Anwendung in der historischen Linguistik. Bachelors thesis, Eberhard Karls Universität Tübingen (2013)
25. Höner zu Siederdissen, C.: Sneaking Around concatMap: Efficient Combinators for Dynamic Programming. In: Proceedings of the 17th ACM SIGPLAN international conference on Functional programming. ICFP '12, New York, NY, USA, ACM (2012) 215–226
26. Höner zu Siederdissen, C., Hofacker, I.L., Stadler, P.F.: How to Multiply Dynamic Programming Algorithms. In: Brazilian Symposium on Bioinformatics (BSB 2013). Volume 8213 of Lecture Notes in Bioinformatics., Springer, Heidelberg (2013) 82–93
27. Höner zu Siederdissen, C., Hofacker, I.L., Stadler, P.F.: Product Grammars for Alignment and Folding. IEEE/ACM Transactions on Computational Biology and Bioinformatics **99**(PrePrints) (2014) 1
28. Meyer, D., Buchta, C.: proxy: Distance and similarity measures. R package version 0.4-3, `http://cran.r-project.org/web/packages/proxy/index.html` (2009)
29. Paradis, E., Claude, J., Strimmer, K.: APE: analyses of phylogenetics and evolution in R language. Bioinformatics **20** (2004) 289–290
30. Saitou, N., Nei, M.: The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular biology and evolution **4**(4) (1987) 406–425
31. Huson, D.H., Richter, D.C., Rausch, C., Dezulian, T., Franz, M., Rupp, R.: Dendroscope: An interactive viewer for large phylogenetic trees. BMC bioinformatics **8**(1) (2007) 460
32. Chen, S.F., Goodman, J.: An empirical study of smoothing techniques for language modeling. In: Proceedings of the 34th annual meeting on Association for Computational Linguistics, Association for Computational Linguistics (1996) 310–318

Named Entity Recognition for User-Generated Content

Sarah Schulz

Ghent University, Groot-Brittanniëlaan 45, 9000 Gent
Sarah.Schulz@UGent.be

Abstract Named entity recognition is a well-studied field of natural language processing. However, for processing of non-standard language, like user-generated content, available tools show a large drop in performance due to the irregularities that can be encountered in such kind of text. In this paper, we investigate a named-entity classifier for user-generated content. We base our experiments on a conditional random fields classifier that has been trained by Desmet and Hoste (2010) on standard Dutch. We show that adding genre-specific features increases performance.

1 Introduction

The Internet holds a variety of possibilities for users to communicate and to express their thoughts and opinions on all sorts of topics. This makes it especially valuable for marketeers and politicians who are interested in an early evaluation of their products or campaigns. Thus, the analysis of such text becomes increasingly important. Tools used in the field of natural language processing are developed to work well for standard language. Analyzing user-generated content (UGC) with common tools shows a significant drop in accuracy (Eisenstein, 2013; Melero et al., 2012). The need to adapt to this kind of data arises.

In this paper, we focus on named entity recognition (NER) which is crucial for a variety of different semantically oriented retrieval tasks like opinion mining, question answering and event tracking. NER is the task of finding names in a text automatically. For the interpretation of opinions uttered in a text, identifying the subjective part and the opinion holder in a sentence is important. NER can either be done using manually compiled pattern rules Farmakiotou et al. (2000) or via the statistical approach of machine learning (Florian et al., 2003; Chieu and Ng, 2002).

We investigate the gain in performance that can be achieve by retraining a classifier on relevant data with respect to UGC. Moreover, we evaluate the influence genre-tailored features can have on the results. We will consider UGC to be a genre following the definition of Trosborg (1997) after whom genres are categories of texts associated by the situation and purpose of their production.

Our research is based on a study performed by Desmet and Hoste (2010). They trained several classifiers for standard Dutch NER and combined them

using ensemble methods. We use their most promising classifier, a conditional random fields classifier, and show the necessity of building a new classifier for UGC. We tackle the task of improving performance for non-standard language in different ways. To show the importance of using UGC for training, we retain the original classifier on our data and compare the results to the results a combined rule-based and gazetteer approach yields. Moreover, we extend the feature set with additional features appearing to be more promising for our dataset. We use a genetic algorithm to perform feature selection and investigate whether this optimization can lead to an increase in performance. The investigation of features suitable for NER in UGC is emphasized in this paper.

In the remainder, we discuss work that has already been conducted in this field. We introduce our dataset and discuss our experimental setup along with the used features. Finally, results are presented and discussed.

2 Related Work

Although a lot of work has been done on NER, the research considering non-standard language is mostly based on the assumption that named entities (NEs) in a specific genre follow easy to generalize patterns. Therefore, many NER systems rely on pattern rules or lexicons. NEs encountered in UGC are partly describable as generalizable depending on the type of social media. In Twitter messages for example there appear a lot of NEs indicated with a hash tag. However, many approaches lack a deeper understanding of the nature of NEs in social media.

Yangarber et al. (2002) describe an approach of NE pattern generalization in the domain of biomedical texts. Such an approach has to our knowledge not been taken for UGC so far. Downey et al. (2007) investigate NER of complex NEs on the web. They use n-gram statistics connected to the theory of multi-word units, but do not account for the nature of UGC. Carvalho et al. (2009) use NE in UGC for irony detection. They work on Portuguese and extract NEs with the help of a NE dictionary. Except for diminutives they also do not consider the deviation from NEs found in standard language. Whitelaw et al. (2008) work on web-scaled NER but rather concentrate on the structure of text found on-line than on the different type of language that can be encountered there.

Gazetteers or approaches making use of rules or patterns, however prove to be not flexible enough as the realization of NEs can highly vary due to abbreviations, spelling and other variation. However, improvements can be achieved through domain adaptation (Daumé and Marcu, 2006). Maynard et al. (2001) consider domain adaptation for NER using different gazetteer lists and grammars. Daumé (2009) describes a domain adaptation approach based on the retraining of a classifier. The approach described makes use of a huge amount of source data associated with another genre and just a small amount of target data

for retraining. Since we have a sufficient amount of UGC available, we decided to focus on retraining a classifier with only target data. Indeed, our focus is not on re-balancing the training set by adding target instances to other genre instances, but on investigating which features might be informative for the retraining of a NER classifier which can handle the peculiarities of UGC.

3 Datasets

In order to retrain a NER system on UGC, it is essential to have gold standard data. In such a gold standard corpus for NER the information whether a token is a NE or not is annotated. We use the dataset of DeClercq et al. (2013) for our experiments.

Table 1: Data statistics of our Dutch corpus representing the amount of data messages, the number of tokens for the three different sub-genres and the amount of NEs.

Sub-genre	**# Messages**	**# Tokens**	**# NEs**
SMS	1,000	16,625	534
SOCIAL NETWORK SITE	1,389	29,817	575
TWITTER	844	12,866	1,706
Total	3,233	59,308	2,815

The dataset has initially been created for the purpose of normalizing UGC. It is compiled of three different sub-genres of UGC, namely social network data, twitter data and text messages. We use the term sub-genre to refer to different subcategories of UGC. We are aware of the problematic nature of the use of this term due to the rather high internal diversity of our sub-genres. It contains about 59,308 tokens out of which 2,815 are marked as NE. The NE mark-up is binary and does not distinguish different sub-types of NEs. The amount of data included in the corpus is shown in Table 1.

Additionally, the corpus contains information about normalization, end of a thought (to account for missing punctuation)[1], regional words, and foreign words. Moreover, words that are ungrammatical[2], stressed, part of a compound

[1] We have deliberately chosen to indicate the ends of thoughts, instead of sentences endings. Annotating sentence endings does not imply it is the end of thought. The author of a text-message could, for example, end sentences with a full stop, and add one or more emoticons. In most cases, however, the emoticon(s) still refer to the previously stated sentence. Consequently, we cannot treat them as separate items.

[2] The category of ungrammatical items is ample and comprises, amongst others, the omission of the subject, the main verb, or the combination of both subject and verb. Moreover, it applies to double negation.

or used as interjections are flagged. However, we only use the original token along with the NE annotation.

Table 2: Example text from all three sub-genres from the corpus showing peculiarities of NEs in UGC.

Sub-genre	**Dutch**	**English**
SMS	Dag **k** en **p**, ons **fie** is geslaagd, oef! Hoe is het daar? Lekker warm zeker?	Hello **k** and **p**, our **fie** passed, phew! How is it there? Certainly nicely warm?
SNS	hey **sarahke** tis al lang gelde dak hier ng op ben geweest ma hey bffl eh ;) x	hey **sarahke** it's been a while since I've been here, but hey bffl eh ;) x
Twitter	**@Mous_tache** drrrrringend! Volgendeweekvrijdag?	**@Mous_tache** urrrrrgent! Nextweek-friday?

Inspection of the data shows that the NEs appearing in such data differ considerably from those appearing in standard data. Short messages often contain strongly abbreviated NEs as shown in Table 2. An example from the social network part of our corpus illustrates the deviation from the standard in terms of dialect use. The "ke"-suffix used to mark diminutive is typical for the Flemish variety of Dutch. Moreover, the use of genre-specific NE characteristics is clarified with the help of an example from twitter data. Hash tags and @-replies are specific ways to mark NEs on certain social network sites and do not appear in standardized texts. It can be noticed that crucial features for NER of standard language in Dutch like capitalization are in most of the cases omitted. The NEs are marked in bold in Table 2.

Moreover, it can be observed that not just the NEs but rather the whole text deviates from the standard. This is due to the fact that there are often limitations of characters on the one hand and that people tend to write as they talk on the other hand. This tendency leads to a graphematic realization different from the one prescribed by orthographic rules. Since classifiers often use context and n-gram information, this can lead to worse results since the same context can look slightly different due to spelling errors, abbreviations, etc.

4 Experimental Setup and Feature Set

In order to investigate the influence of the training data and feature sets on the performance of a classifier, we set different evaluation scenarios. We compared the performance of different classifiers against a gazetteer-and rule-based system as baseline. We used 5-fold cross validation in all our experiments.

4.1 Evaluation Scenarios

(1) Applying a classifier for standard Dutch to our data: Desmet and Hoste (2010) trained classifiers for multi-label classification of six different kinds of NEs

using data from the SoNaR corpus[3] which contains a wide variety of genres of written Dutch. They trained four different classifiers using TiMBL[4]. These classifiers are a memory-based learner, CRF++[5], a conditional random fields classifier (Lafferty et al., 2001), and Yamcha, an open source text chunker oriented toward a lot of natural language processing (NLP) tasks using support vector machines[6]. Moreover, they experimented with different classifier ensembles. The system we used for our experiments is a conditional random fields classifier which showed the best individual performance on their data (F-score 0.82 in average). Different from the data used in the experiments of Desmet and Hoste (2010), our dataset does not contain annotations of different types of NEs like location or person. We aim at binary classification. Since the original models from Desmet and Hoste (2010) have been trained on multi-class labeling we consider everything that is marked as one of the different kinds of NE by the model to be in the positive class.

(2) In a second set of experiments we retrained the classifier on UGC using standard parameter settings and two different feature sets, viz. the feature set described in Desmet and Hoste (2010) and an extended feature set described below.

(3) In a third evaluation scenario, we wanted to investigate the influence of feature selection and retrained the classifier using a genetic algorithm (Desmet et al. 2012) which can perform feature selection for both our feature sets. Genetic algorithms mimic the principle of natural selection. The fitness of each individual in a generations is evaluated. The better it fits the higher the probability to provide the genome for an individual in the next generation. By doing so, fitness improves gradually from generation to generation.
The genetic algorithm *Gallop* allows the optimization of the parameters of a learner with respect to the used features. Individuals are initialized randomly choosing features. The feature selection can be performed on all features separately[7] or on feature groups. The combination of parameters and features compose a genome of a individual. Individuals reaching a high fitness 'survive', individuals with bad fitness are removed from the population. Mutation ensures a diversity of tested genomes. We experimented with rather high mutation probability to lower the probability of converging to local optima. This way the genetic algorithm converges to the best combination of settings.

(4) As a baseline scenario, we implemented a NER based on gazetteer lookup and a pattern rule matching capitalized words. It moreover matches simple patterns like @-replies or hash tags.

[3] `http://taalunieversum.org/archief/taal/technologie/stevin/`, 01/15/2014.
[4] `http://ilk.uvt.nl/timbl/`, 01/15/2014.
[5] `http://crfpp.googlecode.com/svn/trunk/doc/index.html`, 03/04/2014.
[6] `http://chasen.org/~taku/software/yamcha/`, 03/04/2014.
[7] Which is just desirable up to a certain amount of features.

4.2 Feature Sets

Finding the right features is the key to good performance. In the specific case of NER, features have to be extracted which can help to distinguish tokens that are NEs from those that are not. Since we observe a significant drop in performance using the model by Desmet and Hoste (2010) it can be assumed that the features which were used to detect NEs in standard Dutch are less informative for retrieving NEs in UGC. The features used by Desmet and Hoste (2010) are shown in Table 3.

Table 3: Features used for the training of the classifiers by Desmet and Hoste (2010).

#	**Feature Name**	**Data type**	**Description**
			Basic information
1	oriToken	string	the original token
2	POS	string	the part-of-speech (MBSP)
3	first	binary	token in sentence-initial position
			Orthographic information
4	firstCap	binary	first character of token capitalized
5	allCaps	binary	all characters of token capitalized
6	internalCaps	binary	internal characters of token capitalized
7	allLowercase	binary	all characters are lower-cased
8	onlyDigits	binary	token consists only of digits
9	isHypenated	binary	token contains hyphen
10	isPunctuation	binary	token consists of only punctuation marks
11	containsPunct	binary	token contains punctuation marks
12	containsDigit	binary	token contains digits besides other characters
13	containsDigAndAlph	binary	token contains digits besides alpha characters
			Affix information
14	prefix4	string	first 4 characters of token
15	suffix4	string	last 4 characters of token
			Pattern
16	isInitial	binary	token resembles an initial
17	isURL	binary	token is a url
			Other features
18	Word shape	string	symbolic feature which can take the values: allLowercase, allCaps, firstCap, capPeriod, onlyDigits, containsDigitAndAlpha, allCapsAndPunct, firstCapAlphaAndPunct, alphaAndPunct, onlyPunct, mixedCase, other
19	Word length	integer	number of characters
20	Function word	binary	occurs in list of function words
21	Chunks	string	chunk tag (MBSP[8])

Considering the characteristics of UGC as described in Sect. 3, we had to rethink the features. Features like capitalization of the initial letter, word shape, part-of-speech, and chunks which proved to be important in standard text, might be unreliable when handling UGC. Moreover, all the features relying on tools which are trained on standard language such as part-of-speech or chunk information are also problematic as the tags assigned by those tools will further lead to error percolation. The same accounts for the original token itself since there is a higher variation in spelling the same word. Frequently missing orthographic regularity leads to a smaller uniformity of the tokens themselves. Also context information relying on preceding punctuation information is not longer dependable.

Table 4: Additional features for training on UGC.

#	**Feature Name**	**Data type**	**Description**
22	isInGazetteer	binary	token appears in a gazetteer list
23	isInCelex	binary	token appears in Dutch part of the Celex corpus
24	hunspell	binary	uses the open source spell checker Hunspell to check whether token is misspelled or not
25	inverseLength	float	inverse token length
26	prefix3	string	first 3 characters of token
27	suffix3	string	last 3 characters of a token
28	lowerCase	string	token with all characters to lower case

Therefore, in order to improve the performance of the NER, we extended the feature set by the features shown in Table 4 which explicitly account for the characteristics of UGC. The inclusion of a lowercased gazetteer list could help in compensating the missing capitalization. To make sure that NEs which can also be normal Dutch words are not in the gazetteer list, we filtered it using Celex (Baayen et al. 1993). The Celex and Hunspell features check for the possibility of a word being a regular Dutch word which in turn lowers the probability of it being a NE. Celex is a corpus of general lexicons for several languages. This means a token appearing in the Dutch part of Celex has a high probability of being a regular Dutch word. The same accounts for Hunspell. Features 25 to 28 are suggested in Mayfield et al. (2003) for general NER. Since the upper-casing and lower-casing are used also to emphasize emotions in UGC (e.g. I love PETE), lowercasing could help to reach conformity.

5 Results and Discussion

The evaluation and optimization focuses on F-score. Since we rather aim at detecting NEs and not explicitly on classifying each token correctly this is reasonable and makes our results comparable to other research results. We compare

our systems to the gazetteer- and rule-based system baseline, which reaches an F-score of 0.63. Our first classifier, the classifier by Desmet and Hoste (2010), which yielded a 0.82 F-score on standard data, shows a huge drop in performance to an F-score of 0.30 when applied to UGC.

When retraining the NER system of Desmet and Hoste (2010) on our UGC corpus, using their feature set and standard parameter settings, we observe an F-score of 0.87. The same classifier trained on the extended feature set yields a 0.89 F-score. Surprisingly, these results lie above the results reported for standard Dutch by Desmet and Hoste (2010). Inspecting the results, we observed that the good performance we achieve can be related to the high number of NEs following a similar pattern like hash tags or @-replies. Such uniform NEs are easy to recognize for the retrained model. The pattern and rule-based baseline shows that the detection of these specific NEs lead to an F-score of 0.63. Thus in future experiments, we will aim at improving results for the NEs appearing in our corpus which do not belong to this group but are rather inconsistent in spelling and capitalization. This means that features that are considered as reliable for standard language, like capitalization or initial position in a sentence, are probably no longer reliable enough.

Table 5: F-scores achieved by the different classifiers.

Data	**Feature Set**	**Features used**	**F-Score**
standard Dutch	original	1-21	0.30
UGC data	original	1-21	0.87
UGC data	original	1-2,5-8,10-21	0.88
UGC data	extended	1-28	0.89
UGC data	extended	2,3,5,7,8,10-16,18,22,24-28	0.90
Baseline System			
Gazetteer and rule-based approach			0.63

In order to investigate the influence of feature selection, we used a genetic algorithm to optimize the features for our data. We trained a classifier for the original and the extended feature set. The classifier using features selected from the original feature set reaches an F-score of 0.88 which is an noticeable improvement over the classifier trained on standard Dutch and the gazetteer- and rule-based approach (cp. Table 5). It shows a slightly increased performance with respect to the classifier using the complete original feature set although just a small amount of features are deselected:

- #3: first
- #5: allCaps
- #6: internalCaps
- #13: containsDigitAndAlpha

The classifier trained using the genetic algorithm and features selected from the extended feature set reached an F-score of 0.9. This value lies slightly above the result for the classifier using the complete extended feature set and moreover outperforms the optimized classifier trained on the original feature set. The feature selection algorithm identified the features that are important for NER with respect to our dataset.

Expectedly, some features relying on the characteristics of NEs found in standard language, proved to be not important anymore. The following features are not selected by the GA:

- #1: oriToken
- #4: firstCap
- #6: internalCaps
- #9: isHypenated
- #17: isUrl
- #19: wordLength
- #20: function word
- #21: chunks
- #23: isInCelex

This covers with our observations of the differing characteristics of NEs in UGC. Capitalization patterns seem to be not reliable enough to make a decision anymore. Comparing the feature selection results of the original and the extended feature set shows that the capitalization of the first character is just deselected when adding features to replace it. Moreover, the length of a word is not informative. This could be due to the fact that NEs in UGC vary from abbreviations to really long strings like @-replies. An explanation for the omission of the chunk feature could be the fact that it relies on a tagging task. Since UGC is hard to process with NLP tools, the chunk tags could often be unreliable and therefore not be informative. Surprisingly, the newly added feature telling whether a word appears in Celex or not, seems to be uninformative.

The results for the different scenarios are summed up in Table 5.

6 Conclusion and Future Work

We investigated the influence of training data and feature selection on the performance of NER in the context of UGC.

We show that retraining on the relevant genre outperforms a rather naive approach based on gazetteer lookup and capitalization and a learning approach trained on standard data. This is not surprising since in-domain training and testing is an easier task than cross-genres classification. More interestingly, we could show an improvement by adding genre-tailored features. Retraining the model with the extended feature set in combination with feature selection improves the results. In sum, we improved results from an F-score of 0.30 to 0.90 by retraining the classifier in combination with optimizing the used features. The selected features mirror the specificities of named entities in UGC.

We show that genre-tailored features can have an effect on performance, although we could not show a strong increase in performance. This general trend of increase allows space for further improvement of the used feature set which requires an even further analysis of the specifics of NE in UGC.

So far, we are working with a rather small dataset which is sufficient to get an impression of which features can be helpful for NER in UGC. Results could be improved and especially made more transferable by increasing the size of our training data. This could be done using the approach of domain adaptation. Alternatively, self-training as a manner of extending the training data in an unsupervised way could be promising as well. Moreover, we show that genre can influence the required features. Therefore, it seems promising to include more sub-genres of UGC into the training in order to make the system more robust with respect to genre specifics.

References

1. Baayen, R., Piepenbrock, R., and van Rijn, H.: The CELEX lexical database on CD-ROM (2013)
2. Carvalho, P., Sarmento, L., Silva, M. J., and de Oliveira, E.: Clues for detecting irony in user-generated contents: Oh...!! it's "so easy" ;-). In Proceedings of the 1st International CIKM Workshop on Topic-sentiment Analysis for Mass Opinion (TSA '09), 53–56 (2009)
3. Chieu, H. L. and Ng, H. T.: Named entity recognition: A maximum entropy approach using global information. In Proceedings of the 19th International Conference on Computational Linguistics 1, (COLING '02), 1–7, (2002)
4. Daelemans, W., Buchholz, S., and Veenstra, J.: Memory-based shallow parsing. In CoNLL-99, 53–60 (1999)
5. Daumé,III, H.: Frustratingly Easy Domain Adaptation. Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, 256–263 (2007)
6. Daumé,III, H. and Marcu, D.: Domain Adaptation for Statistical Classifiers. J. Artif. Int. Res. 26/1, 01–126 (2006)
7. De Clercq, O. and Schulz, S. and Desmet, B. and Lefever, E. and Hoste, V.: Normalization of Dutch User-Generated Content. Proceedings of the 9th International Conference on Recent Advances in Natural Language Processing (RANLP 2013), 179–188 (2013)
8. Desmet, B. and Hoste, V.: Dutch Named Entity Recognition using Classifier Ensembles. Computational Linguistics in the Netherlands 2010: selected papers from the twentieth CLIN meeting, 29–41 (2010)
9. Desmet, B. and Hoste, V. and Verstraeten, D. and Verhasselt, J.: Gallop Documentation. Technical Report Language And Translation Technology Team, Ghent University (2012)
10. Downey, D., Broadhead, M., and Etzioni, O.: Locating complex named entities in web text. In In Proc. of IJCAI, 2733–2739 (2007)
11. Eisenstein, J.: What to do about bad language on the internet. HLT-NAACL, 359–369 (2013)
12. Farmakiotou, D., Karkaletsis, V., Koutsias, J., Sigletos, G., Spyropoulos, C. D., and Stamatopoulos, P.: Rule-based named entity recognition for greek financial texts. In Proceedings of the Workshop on Computational Lexicography and Multimedia Dictionaries (COMLEX 2000), 75–78 (2000)
13. Florian, R., Ittycheriah, A., Jing, H., and Zhang, T.: Named entity recognition through classifier combination. In Proceedings of the Seventh Conference on NaturalLanguage Learning at HLT-NAACL 2003/4 (CONLL '03), 168–171 (2003).
14. Lafferty, John D. and McCallum, Andrew and Pereira, Fernando C. N.: Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. Proceedings of the Eighteenth International Conference on Machine Learning (ICML 2001), 282–289 (2001)
15. Mayfield, J., McNamee, P., and Piatko, C.: Named entity recognition using hundreds of thousands of features. In Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003/4 (CONLL '03), 184–187, (2003)
16. Maynard, D., Tablan, V., Ursu, C., Cunningham, H., and Wilks, Y.: Named entity recognition from diverse text types. In Proceedings of the Recent Advances in Natural Language Processing 2001 Conference,257–274 (2001)

17. Melero, M., Costa-Juss'a, M. R., Domingo, J., Marquina, M., and Quixal, M.: Holaaa!! writin like u talk is kewl but kinda hard 4 nlp. In Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12), 3794–3800 (2012).
18. Trosborg, A.: Text Typology and Translation. J. Benjamins, Amsterdam, Netherlands (1997)
19. Whitelaw, C., Kehlenbeck, A., Petrovic, N., and Ungar, L.: Web-scale named entity recognition. In Proceedings of the 17th ACM Conference on Information and Knowledge Management (CIKM '08), 123–132, (2008)
20. Yangarber, R., Lin, W., and Grishman, R.: Unsupervised learning of generalized names. In Proceedings of the 19th International Conference on Computational Linguistics 1, (COLING '02), 1–7, (2002)

A Corpus-Based Approach to English Modal Adverbs in the Study of Synonymy

Daisuke Suzuki

Kyoto University, Japan
suzuki0213@gmail.com

Abstract This study examines English modal adverbs in the study of synonymy by focusing on the two pairs of synonymous modal adverbs: *doubtless*, *no doubt* and *maybe*, *perhaps*. After extracting data regarding these adverbs from the British National Corpus (BNC), this study determines three factors by analyzing the target adverbs in a larger context: (i) the co-occurrence of the target adverbs with modal verbs, (ii) the position (i.e., initial, medial, or final) the target adverbs occupy in a clause, and (iii) which pronouns fill the subject slot in their clauses. The results of the analysis demonstrate that the modal adverbs fulfill different functions at the discourse-pragmatic level, and the factors influencing the use of these modal adverbs are strongly associated with the parameters of modality, discourse, and interaction. These results can be applied to establish clear usage guidelines for the adverbs.

Keywords: synonymy, modal adverbs, functional analysis, corpus data

1 Introduction

Synonymy is one of the best-known semantic relations between lexical items. The word "synonym" customarily denotes paired items that either share a meaning or convey very similar meanings. However, in terms of language function and use, it is impossible for synonyms to be mutually interchangeable in all environments. Originally, the fundamental assumption was that "if the forms are phonemically different, we suppose that their meanings also are different," that is, "there are no actual synonyms" (Bloomfield 1933: 145).[1] Indeed, Lyons (1968: 448) and Palmer (1981: 89) agree that there are very few "real" or "perfect" synonyms in a language. Hence, "one would expect either that one of the items would fall into obsolescence or that a difference in semantic function would develop" (Cruse 1986: 270). Viewed in this light, synonyms are considered to function differently in actual usage. Thus, I can clearly distinguish meaning between a given pair of synonyms by considering their functions in the context of a clause or discourse.[2]

[1] For this reason, I use the term "synonym" in this paper to refer to an example of a set that has a semantic relation of "near synonymy."

[2] The term "discourse" is intended to underline the fact that these adverbs must be described at the level of discourse rather than at the sentence level.

English Modal adverbs are also rich in synonyms. In the present study, I focus on the two pairs of synonymous modal adverbs: *doubtless*, *no doubt* and *maybe*, *perhaps*. These are similar in form and nearly equivalent in meaning in each pair, and thus classified in the same semantic category. As demonstrated by Examples (1–2), these modal adverbs are used to express a speaker's judgment regarding the probability or possibility of a proposition:

(1) You have *doubtless* or *no doubt* heard the news. (Fowler 1998: 230)
(2) *Maybe/Perhaps* it'll stop raining soon. (Swan 2005: 348)

Despite their similarity in form and meaning, I conduct a detailed analysis to distinguish them on the basis of corpus data. As Biber, Conrad and Reppen (1998: 24) mentioned, investigating the use and distribution of synonyms in a corpus enables us to determine their contextual preferences. I identify what factors are significant in predicting each adverb's usage and how these adverbs differ. Moreover, by identifying the functional distinctions between each pair of synonyms, I provide a foundation from which to develop clear guidelines for their usage.

2 Previous studies

According to *Merriam-Webster's Dictionary of English Usage*, *doubtless* and *no doubt* imply some doubt and are used to mean "(very) probably" despite their denotative word-formations (p. 369). Quirk et al. (1985: 623), Fowler (2004: 230), and Swan (2005: 378) also propose this description. With regard to *maybe* and *perhaps*, these are nearly synonymous, following *Longman Language Activator* and *Oxford Thesaurus of English*.

Compared to the increasing number of studies on *no doubt* (Quirk et al. 1985: 623; Biber et al. 1999: 854; Huddleston and Pullum 2002: 768; Fowler 2004: 230; Swan 2005: 378; Simon-Vandenbergen and Aijmer 2007) and *perhaps* (Greenbaum 1969: 153; Bellert 1977: 344; Lyons 1977: 798; Perkins 1983: 89–92, 101–104; Watts 1984: 137–138; Quirk et al. 1985: 620; Doherty 1987: 53; Swan 1988: 459–460; Biber et al. 1999: 854; Huddleston and Pullum 2002: 768; Swan 2005: 348; Tancredi 2007: 2; Ernst 2009: 515), there have been relatively less studies on *doubtless* and *maybe*, and the existing literature offers no clear means of determining how and when each adverb is likely to be used within a particular construction or context. In order to investigate the pragmatic characteristics of these modal adverbs, I perform a functional analysis that provides new insights into the behaviors of these modal adverbs.

3 Methodology

The data adduced in this study to conduct a functional analysis of the modal adverbs are from the British National Corpus (BNC) because its large scale and wide range of genres provide sufficient data concerning the use of the four modal

adverbs for various purposes within diverse contexts. The BNC, a 100-million-word corpus, includes both written (90%) and spoken (10%) British English. To prepare the data for analysis, I first extracted all occurrences of the target adverbs from the corpus. I then examined each occurrence manually to identify those in which one of the four modal adverbs functioned as a sentence adverb;[3] I identified **731** such instances of *doubtless*, **2,701** of *no doubt*, **6,694** of *maybe* and **22,189** of *perhaps*. A quantitative analysis of these findings was also conducted to test for frequency.

In my analysis of the modal adverbs, I focused on the larger context in which those expressions occurred, and I investigated three factors regarding their patterns of occurrence: (i) which modal verbs they co-occurred with,[4] (ii) in which position of the three (i.e., initial, medial, or final) they occurred in a clause,[5] and (iii) which pronouns filled the Subject slot in their clauses. In order to calculate the frequency of occurrence in terms of position and function, I determined the frequency of each adverb in each position as well as the frequency of co-occurrence with modal verbs and subject pronouns. I then examined the number of occurrences per thousand of these modal adverbs with each of the modal verbs and subject pronouns in the BNC. Thus, I explored the use of *doubtless*, *no doubt*, *maybe*, and *perhaps* in the modal, discursive, and interpersonal contexts by examining (i), (ii), and (iii), respectively, as defined above.

---

[3] For this analysis, I excluded all examples of one-word utterances for response, such as "*No doubt.*" and "*Perhaps.*" Also excluded were examples that did not form a complete clause, such as "*Perhaps* only in the next life." (BNC: A08). In addition, I excluded examples where the modal adverbs occurred within the phrase structure (i), and where they modified not a clause but a phrase in which a comma (**,**) intensified the expressed meaning (ii), as in the following:

(i) I haven't been to an organized campsite for *perhaps* fifteen years, so all this is new to me. (BNC: A6T)
(ii) One snap even shows him on top of her**,** *no doubt* for closer inspection. (BNC: CH5)

[4] I confined my focus to the nine modal verbs, *can*, *could*, *may*, *might*, *shall*, *should*, *will/'ll*, *would/'d*, *must*, which Quirk et al. (1985: 137) and Biber et al. (1999: 73) classify as "central modal auxiliaries".

[5] In Quirk et al. (1985: 490–491) and Hoye (1997: 148), the positions in which they appear are presented as follows:

(a) *Possibly* they may have been sent to London. [initial]
(b) They *possibly* may have been sent to London. [initial-medial]
(c) They may *possibly* have been sent to London. [medial]
(d) They may have *possibly* been sent to London. [medial-medial]
(e) They may have been *possibly* sent to London. [end-medial]
(f) They may have been sent *possibly* to London. [initial-end]
(g) They may have been sent to London *possibly.* [end] (Hoye 1997: 148)

4 Results and discussion

First, I examined the possibility of their co-occurrence with modal verbs. Tables 1 and 2 show the frequency and percentage with which each modal adverb co-occurred with the modal verbs in the BNC. As shown in Tables 1 and 2, *doubtless* and *maybe* tends to co-occur more frequently with modal verbs than *no doubt* and *perhaps*, respectively.

Table 1: Frequency and percentage of co-occurrence with modal verbs

Modal adverb	**Total**	**Freq.**	**%**
doubtless	731	327	**44.7**
no doubt	2,701	1,065	39.4

($\chi^2 = 6.49$, d. f. $= 1$, $p < 0.05$)

Table 2: Frequency and percentage of co-occurrence with modal verbs

Modal adverb	**Total**	**Freq.**	**%**
maybe	6,694	2,317	**34.6**
perhaps	22,189	6,552	29.5

($\chi^2 = 62.25$, d. f. $= 1$, $p < 0.001$)

The collocation with modal verbs suggests that the target expression implies modality, that is, the speaker expresses his or her mental attitude toward the proposition. Thus, the use of *doubtless* and *maybe* is likely a means of reinforcing the expression of modality. This function is illustrated in Examples (3–6):

(3) All-time greatness **would** *doubtless* be bestowed upon Carling, already a veteran captain at 26. (BNC: K4T)
(4) *No doubt* they**'ll** find Dad and Pet before long. (BNC: AN7)
(5) *Maybe* Francis **will** think of me kindly one day. (BNC: CDY)
(6) The stronger fish **will** *perhaps* reach 4 lb. (BNC: B0P)

Figure 1 illustrates the co-occurrence patterns between the target adverbs and modal verbs.[6] Because of the variety of the types of modal verbs and the differences among their occurrence with the target adverbs, it is difficult to directly identify a clear-cut trend in the usage of these adverbs. For this reason, I

[6] Data pertaining to Figures 1 and 2 are provided in the appendix.

employed a statistical technique referred to as correspondence analysis (CA).[7] As shown in Figure 1, when two of the row and column variables are plotted at a relatively close range, we can identify a strong affinity or close association between them. Thus, *maybe* is seen as closely correlated with *would* and *will*, whereas *perhaps* is more frequently correlated with *can* and *could*. Moreover, *perhaps* lies in the same quadrant as *may* and *might*, which express low probability. These findings concerning the relationship between modal verbs and adverbs indicate that *maybe* implies a higher possibility than *perhaps* in terms of likelihood. As observed, on the other hand, we can identify an association between *doubtless* and *no doubt*, indicating that the two adverbs conveys nearly the same degree of probability.

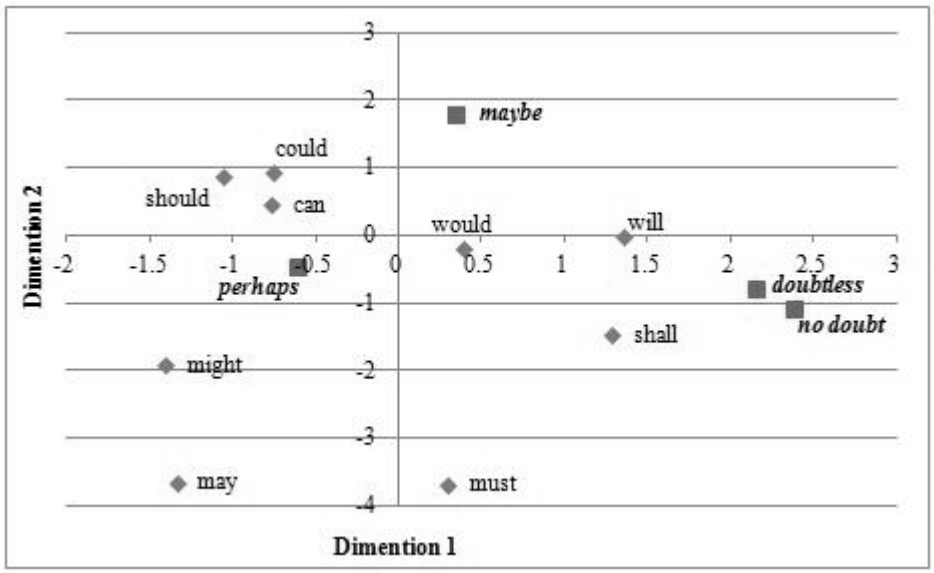


Figure 1: Results of CA of the target adverbs and modal verbs in the BNC

Turning to examination of the position, Figure 2 illustrates the percentage of total occurrences in which the modal adverbs are positioned in the initial, medial, and final position. Illustrations of each position are shown in Examples (7–14). We can observe that the initial use of *no doubt* and *maybe* is strongly

[7] Developed by Benzécri in the 1960s, CA is one of the multivariate techniques used to summarize information regarding multivariate data. Alongside principal components analysis (PCA) and factor analysis (FA), CA is used to analyze grouped objects and variables and provide a graphic display of the results. In CA, all the row and column coordinates are simultaneously given quantities so that a correlation coefficient between the row and column coordinates can be maximized.

preferred in each pair; particularly, the initial *maybe* is markedly high in the BNC.

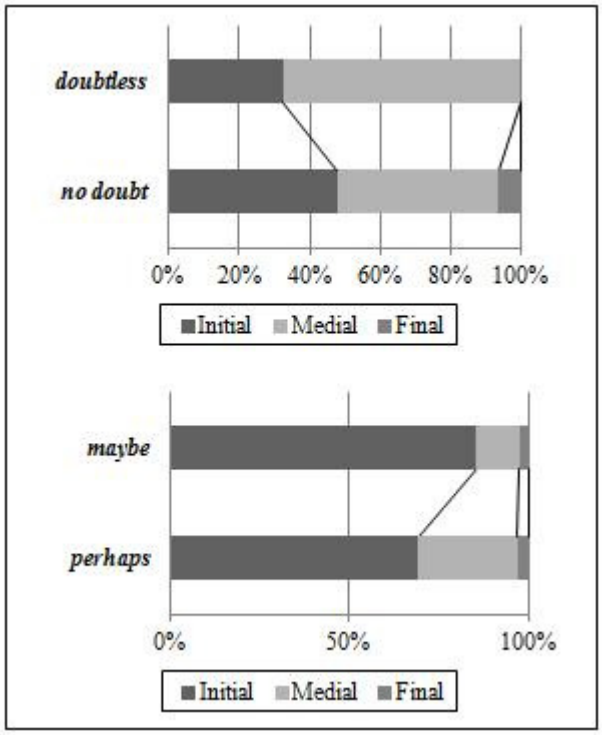


Figure 2: Proportions of the positions of the target adverbs

(7) *Doubtless* there were many occasions night and day when a tempest was raging outside. (BNC: K7E)
(8) The ascent was *doubtless* relatively easy. (BNC: EFR)
(9) *No doubt* some of them volunteered for war service a year later. (BNC: B1P)
(10) Pupils will at first *no doubt* compare and contrast the past and the present. (BNC: HXF)
(11) *Maybe* she would even smell a whiff of perfume. (BNC: AC3)
(12) I often thought it was *maybe* for the birds. (BNC: H7A)
(13) *Perhaps* she would enjoy it more if Lydia stopped whimpering. (BNC: G0X)
(14) The persistence of elections is *perhaps* odd. (BNC: CR9)

As Halliday (1970: 335), Perkins (1983: 102–104), Hoye (1997: 148–152), and Halliday and Matthiessen (2004: 79–85) observe, a modal adverb positioned initially expresses the topic or theme of modality, as illustrated in Examples (15) and (16):

(15) *Possibly* it was Wren.
(16) It *may* have been Wren. (Halliday 1970: 335)

Examples (15) and (16) convey the same meaning in terms of probability; however, the use of *possibly* in Example (15) fulfills the discourse function of expressing the topic or theme. In this sense, the modal adverbs occurring initially can serve as guidelines for the hearer or reader regarding the flow of discourse; in other words, they have the pragmatic function of structuring the discourse. Hence, this analysis indicates that *no doubt* is strongly attracted to, and *maybe* is restricted to, the discourse function of topic encoding in actual use.

Moving on to the final major finding in the BNC is related to the frequency of the modal adverbs' co-occurrence with subject pronouns. As shown in Tables 3 and 4, which illustrate the co-occurrence patterns among the modal adverbs and *I*, *you*, and *we*, *no doubt* and *maybe* is used more frequently with first and second person pronouns than *doubtless* and *perhaps*, respectively.

Table 3: Frequency and percentage of co-occurrence with person pronouns

	doubtless **(731 instances)**		*no doubt* **(2,701 instances)**	
Subject pronoun	**Freq.**	**Per 1,000**	**Freq.**	**Per 1,000**
I	9	12.3	49	18.1
you	27	36.9	182	67.4
we	4	5.5	55	20.4

Table 4: Frequency and percentage of co-occurrence with person pronouns

	maybe **(6,694 instances)**		*perhaps* **(22,189 instances)**	
Subject pronoun	**Freq.**	**Per 1,000**	**Freq.**	**Per 1,000**
I	849	126.8	1,610	72.6
you	682	101.9	1,764	79.5
we	512	76.5	1,216	54.8

This fact is illustrated in Examples (17) and (18), in which, as clearly shown by co-occurrences with the first or second person pronoun (and in interrogatives), we can observe occurrences of interpersonal uses of *no doubt* and *maybe*. These pronouns linguistically express people concerned in a conversation in an explicit way. Thus, the use of *no doubt* and *maybe*, in contrast to that of *doubtless* and *perhaps*, respectively, is preferred in the context involving the speaker and hearer.

(17) *No doubt*, **you** would do anything to catch her murderer? (BNC: ANL)
(18) *Maybe* **we** should go to one of the hotels for tea or icecream? (BNC: A6N)

Moreover, the marked pattern of the use in interrogative contexts can be observed in the BNC. The following are examples of *no doubt* and *maybe*, used as meta-linguistic devices to confirm or emphasize information and understanding in the interactive process involving the speaker and hearer—that is, to fulfil an interpersonal function in the conversation. There is an interesting shift regarding the use of *no doubt* and *maybe* from expressing the speaker's mental attitude to marking shared familiarity with the hearer.

(19) **You** have heard different versions, *no doubt*? (BNC: G1A)
(20) **You** have read it, *no doubt*? (BNC: GVP)
(21) **You** wouldn't recognise us with our clothes on, *maybe*? (BNC: HTS)
(22) Do **you** do the same line of work, *maybe*? (BNC: FPM)

In terms of modality, *doubtless* is strongly associated with the modal function; in terms of discourse, *no doubt* is strongly associated with expression of the topic or theme in a clause and interpersonal uses. On the other hand, *maybe* is closely related to all of the three functions.

5 Conclusion

This study investigated whether the usage of English modal adverbs was associated with the discourse-pragmatic context in which they occurred. To explore the factors determining their usage within the broader contexts in which they occurred, I analyzed data extracted from the BNC corpus, which provided usage data for *doubtless*, *no doubt*, *maybe*, and *perhaps* in natural settings. The modal adverbs that at first sight appear to be exchangeable in a variety of contexts can be distinguished on the basis of their detailed functional characteristics.

Examining the functions of the four adverbs from the modal, discourse, and interpersonal points of view, I demonstrated that these modal adverbs fulfilled different functions at the discourse-pragmatic level. I elucidated that factors related to discourse-pragmatic domain, such as those examined in this study, were generally valid in the study of synonymy. These findings suggest that in synonym study, we need to examine a wider range and level of factors that can influence the choice in a synonym pair. Finally, these fine-grained distinctions between the synonymous modal adverbs provided a significant foundation for the comparison of these modal adverbs, which is necessary in the development of clear guidelines for their usage.

Appendix: Data for Figures 1–2

Data for Figure 1

Modal adverb	**Initial**	**Medial**	**Final**	**Total**
doubtless	237	492	2	731
no doubt	1,288	1,237	176	2,701
maybe	5,725	808	161	6,694
perhaps	15,334	6,179	676	22,189

Data for Figure 2

Modal verb	*doubtless*	*no doubt*	*maybe*	*perhaps*
must	3	9	5	39
will	158	594	701	1,222
would	131	330	581	1,510
shal	3	25	20	49
should	4	9	331	1,082
can	13	34	228	746
could	13	50	369	1,088
may	1	11	16	305
might	1	3	66	512
Total	327	1,065	2,317	6,553

References

1. Bellert, I.: On semantic and distributional properties of sentential adverbs. Linguistic Inquiry 8(2), 337–351 (1977)
2. Biber, D., Johansson, S., Leech, G., Conrad, S.: Longman Grammar of Spoken and Written English. Pearson, Harlow (1999)
3. Biber, D., Conrad, S., Reppen, R.: Corpus Linguistics: Investigating Language Structure and Use. Cambridge University Press, Cambridge (1998)
4. Bloomfield, L.: Language. Holt, New York (1933)
5. Cruse, D.A.: Lexical Semantics. Cambridge University Press, Cambridge (1986)
6. Doherty, M.: Perhaps. Folia Linguistica 21(1), 45–65 (1987)
7. Ernst, T.: Speaker-oriented adverbs. Nat Lang Linguist Theory 27, 497–544 (2009)
8. Fowler, H.W.: Fowler's Modern English Usage, 3rd edn. Revised by Burchfield, R.W. Oxford University Press, Oxford (2004)
9. Greenbaum, S.: Studies in English Adverbial Usage. University of Miami Press, Coral Gables (1969)
10. Halliday, M.A.K., Matthiessen, C.M.I.M.: An Introduction to Functional Grammar, 3rd edn. Arnold, London (2004)
11. Halliday, M.A.K.: Functional diversity in language as seen from a consideration of modality and mood in English. Foundations of Language 6, 322–361 (1970)
12. Hoye, L.: Adverbs and Modality in English. Longman, London (1997)
13. Huddleston, R., Pullum, G.K.: The Cambridge Grammar of the English Language. Cambridge University Press, Cambridge (2002)
14. Longman Language Activator, 2nd edn. Pearson Education, Harlow (2002)
15. Lyons, J.: Introduction to Theoretical Linguistics. Cambridge University Press, Cambridge (1968)
16. Lyons, J.: Semantics 2. Cambridge University Press, Cambridge (1977)
17. Merriam-Webster's Dictionary of English Usage. Merriam-Webster, Massachusetts (1994)
18. Oxford Learner's Thesaurus: A Dictionary of Synonyms. Oxford University Press, Oxford (2008)
19. Oxford Thesaurus of English, 3rd edn. Oxford University Press, Oxford (2009)
20. Palmer, F.R.: Semantics, 2nd edn. Cambridge University Press, Cambridge (1981)
21. Perkins, M.R.: Modal Expressions in English. Frances Pinter, London (1983)
22. Quirk, R., Greenbaum, S., Leech, G., Svartvik, J.: A Comprehensive Grammar of the English Language. Longman, London (1985)
23. Simon-Vandenbergen, A.: No doubt and related expressions: A functional account. In: Hannay, M., Steen, G.J. (eds.) Structural-Functional Studies in English Grammar: In Honour of Lachlan Mackenzie, pp. 9–34. John Benjamins, Amsterdam (2007)
24. Simon-Vandenbergen, A., Aijmer, K.: The Semantic Field of Modal Certainty: A Corpus-Based Study of English Adverbs. Mouton de Gruyter, Berlin (2007)
25. Swan, M.: Practical English Usage, 3rd edn. Oxford University Press, Oxford (2005)
26. Swan, T.: Sentence Adverbials in English: A Synchronic and Diachronic Investigation. Novus, Oslo (1988)
27. Tancredi, C.: A multi-model modal theory of I-semantics: Part I: modals. Ms. University of Tokyo (2007)
28. Watts, R.J.: An analysis of epistemic possibility and probability. English Studies 65(2), 129–140 (1984)

Slash/A N-gram Tendency Viewer – Visual Exploration of N-gram Frequencies in Correspondence Corpora*

Velislava Todorova† and Maria Chinkina

University of Tübingen

1 Introduction

In this paper we present a visualization web tool which we have developed for the analysis of tendencies in the change of language over time. Slash/A N-gram Tendency Viewer[1] (simply *Slash/A* from now on) is designed for the exploration of n-gram frequencies in correspondence corpora. It represents the frequencies of selected n-grams as a graph in a coordinate system with time on the x axis and frequency on the y axis. Slash/A also provides the option of smoothing the graph, making the general tendency clearer to see. Smoothing eliminates (or at least limits) possible sources of confusion, like exceptional extreme values or overlaps when multiple graphs are presented.

We will explain how we process data and what linguistic information we extract from it. We will also discuss the visualization techniques which we used for the representation of this information.

2 Application

Slash/A is built to facilitate the discovery and exploration of dependencies between linguistic elements and of patterns in language use over time.

For example, querying the second volume of the Brownings' corpus,[2] which we used as our development corpus, we found some interesting correlations. This

* We are grateful to Dr. Christopher Culy, who supervised our work on the project and revised several versions of this paper; to Plamen Trayanov for the valuable discussion on the smoothing techniques and to the three anonymous reviewers for their comments and suggestions.

† I am also thankful to the DAAD for supporting my studies in the University of Tübingen where this paper was written.

[1] The name of the tool comes from the names of its authors: Slava and Masha, and (quite ironically) is connected to the only sign it can not process, the slash, which functions as a separator between the elements of a query, separating for example the token from its POS tag. More about the syntax of the queries can be found in section 4.

[2] The corpus is annotated and formatted in TCF format, and is available from `http://www.sfs.uni-tuebingen.de/~cculy/vistola/#resources`.

corpus consists of love letters exchanged between Robert Browning and Elizabeth Barrett over a period of almost two years, after which they got married. We compared the frequencies of the words *love* and *happy*. After smoothing the results, one can see (Fig. 1) that most of the time the frequencies of these words increase and decrease together, except for the last couple of weeks when the usage of *love* goes up, while the opposite happens with *happy*. One explanation might be that the prospect of the upcoming marriage increased the use of *love*, while the disapproval of Elizabeth's father resulted in decrease of the use of the word *happy*. Figure 2 shows that the reference to Elizabeth's father by both correspondents is more often whenever the topic of marriage is discussed, it also suggests that Robert was more concerned with the issue.

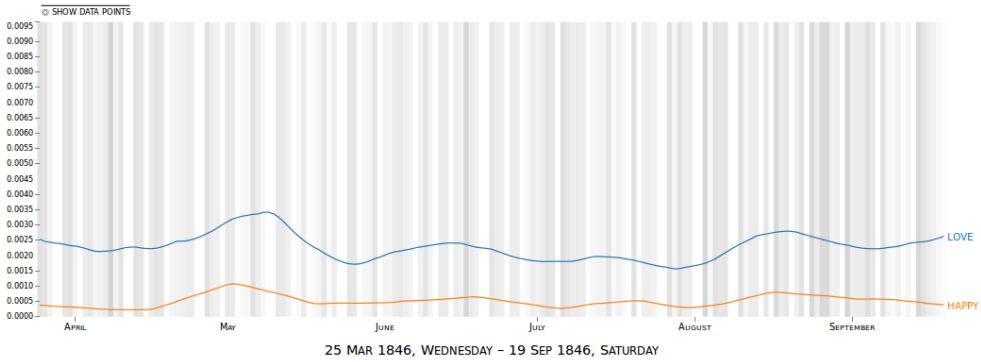


Figure 1: Frequencies of the words *love* and *happy* in the Brownings' corpus in the period between March 25, 1946 and September 19, 1946.

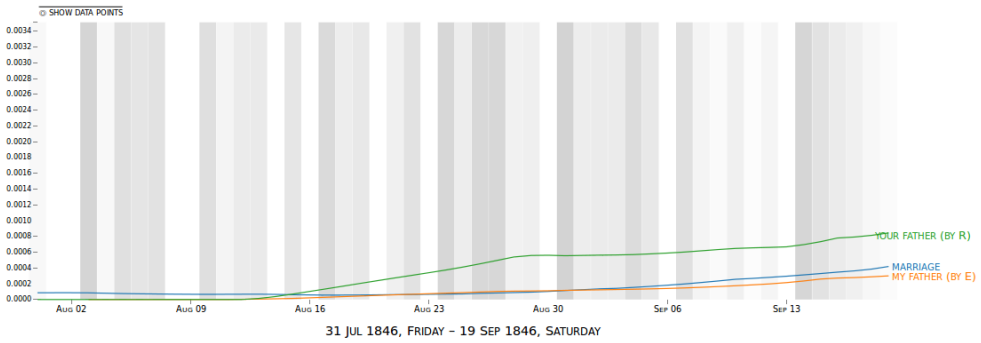


Figure 2: Frequencies of the n-grams *marriage* (as used both by the two authors), *my father* (only in Elizabeth's letters) and *your father* (only in Robert's letters) in the period between July 31, 1946 and September 19, 1946.

There are other tasks Slash/A can be used for, as for example author or date identification. (The annotated and proofread Brownings' corpus itself contains 32 letters with missing date and/or author in the metadata.) The task of topic detection could also make use of the tool.

3 Comparable tools

Slash/A resembles Google Ngram Viewer[3], but there are some important differences that need to be mentioned. First of all, Google Ngram Viewer is visualizing information that has already been extracted from a fixed corpus, while Slash/A takes as an input a user specified text corpus and conducts all the necessary searches on the go. This has several noteworthy implications. Most importantly, our tool is very convenient for researchers interested in particular collections of texts and not in the content of the Google Books corpus[4]. Besides, the user can search for n-grams of any length. Google Ngram Viewer cannot display sequences of more than 5 tokens, because the preliminary search was restricted to five-grams. From a technical point of view, Slash/A is a simpler tool, because it does not need a component handling lists of n-grams obtained after searching the corpus for them.

Further, Slash/A has some additional features. It allows filtering by author; the user has a direct access to the original text and we let the user specify their own smoothing parameter. Moreover, our smoothing algorithm uses a weighted moving average instead of a simple one which ensures less angular view.

At the end, there are several functionalities of the Google viewer that we haven't implemented in Slash/A. One of them is the option to switch between case sensitive and case insensitive mode, which is something we are looking forward to introduce in our tool too. Another useful feature is the possibility to combine multiple time series into one. There are also the n-gram subtraction and multiplication and the very specific use of wildcards, allowing the user to see the top ten examples for n-grams of a given form.

4 N-gram queries

The n-grams that Slash/A works with are sequences of n tokens and there is no limit for their length.

The tool accesses the annotations for tokens, lemmas and POS tags and all of them can be used most creatively for the composition of a corpus query. The following are examples of valid queries using the Penn Tree Bank tag set:[5]

[3] The viewer is available from `https://books.google.com/ngrams#`; a detailed description can be found on `https://books.google.com/ngrams/info`.

[4] `http://books.google.com/`

[5] The PTB POS tag set can be found here: `http://www.cis.upenn.edu/$\sim$treebank/`

my book	query for the bi-gram *my book*
my book/lemma	query for all bi-grams with first element *my* and second element any form of the word *book*
/VBP book/NN	query for all bi-grams with first element a verb in non-third person singular present tense and second element the singular form of the noun *book*
/V* book	query for all bi-grams with first element a verb and second element *book*

The last example illustrates the way we allow the use of a wildcard character (*) for POS tags. If the user only provides the first letter(s) of the tag followed by an asterisk, all the tags starting with this (sequence of) letter(s) will be matched.

There is no wildcard that can be used with tokens. However the third example shows that omitted token in the query leads to matching any token with the specified POS tag.

5 Smoothing

After searching the corpus for a particular n-gram, we obtain a number of data points - information about the frequency of this n-gram at a point of time. If the corpus is such that there is no data at certain points of time, we linearly interpolate frequency values for these points.

The resulting graph rarely allows the eye to see the general tendency of the frequency change behind the multiple ups and downs. Only small corpora and n-grams with particularly steadily changing frequencies produce an easy to trace curve. For the general case some sort of smoothing is desirable.

We use a linearly weighted moving average to smooth the frequency graph. We decided to use days as base time units. Let D be the number of days in the time period covered by the corpus. We calculate the smoothed frequency value s_d for the d-th day of the period ($1 \leq d \leq D$) with the following formula:

$$s_d = \frac{\sum_{i=d-p}^{d+p}(p - |d-i| + 1)w_i f_i}{\sum_{i=d-p}^{d+p}(p - |d-i| + 1)w_i} \tag{1}$$

f_i is the frequency for the i-th day. If $1 \leq i \leq D$ and there is no data in the corpus for the i-th day, this value is obtained by linear interpolation. If $i < 1$ or $i > D$, we can assume that $f_i = 0$ (in this case the weight $w_i = 0$, which renders the frequency value irrelevant).

p is the smoothing parameter, i.e. it determines the size of the averaging window or, to put it differently, the number of days which are taken into account when calculating the smoothed frequency value for the d-th day. The size of the averaging window is $2p + 1$, namely p days in the past, p days in the future and the d-th day itself. The set of values for p that we take for our predefined levels of smoothing is $\{3, 15, 45, 182, 1825\}$, which corresponds to the following set of time periods: $\{week, month, trimester, year, decade\}$. We also allow the user to specify their own smoothing parameter.

w_i is the weight of f_i. These weights are calculated on the basis of the overall number of tokens T_i written on the i-th day, which ensures high accuracy of the results, by giving little weight to observations based on little evidence, as they are likely to be noisy. We spread weights from the data points to the time points for which there is no data in the corpus, to avoid zero weights in time points from the period of interest.

If $i < 1$ or $i > D$, $w_i = 0$; otherwise it is calculated as follows:

$$w_i = \begin{cases} \sqrt{T_i} & \text{if } T_i > 0 \\ \dfrac{w_l}{i - l + 1} + \dfrac{w_r}{r - i + 1} & \text{if } T_i = 0. \end{cases} \tag{2}$$

l and r are day indices. The l-th day is the closest day to the i-th in the past (i.e. to the left on the time axis), such that there is data in the corpus for it. Formally, l is such that $T_l > 0$ and $l < i$ and if for some x $x < i$, then $x \leq l$. Respectively, r refers to the closest data point to i in the future (i.e. to the right on the time axis) and, put formally, r is such that $T_r > 0$ and $r > i$ and if for some x $x > i$, then $x \geq r$. $(i - l)$ and $(r - i)$ are the distances to the closest data point in the past and in the future respectively. The greater the distance between a no-data point to the closest data points, the smaller the portion of their weights that we assign to this data point.

The expression $(p - |d - i| + 1)$ in (1) is a second type of weight. When the smoothed frequency value s_i is calculated, the values for the days closer to the i-th are taken as more significant, the values close to the edges of the averaging window have less effect on s_i. These weights are needed to obtain a curve that intuitively can be called smoother than the original graph, i.e. a curve that changes its direction less and forms less visible angles or at least less acute angles.

We have tried alternative techniques, but the results were unsatisfactory. With dense corpora (with data for almost every day in the period) the observed differences were smaller, sparse data on the other hand seemed to be more problematic because of the big gaps between data points. The parts of the graph corresponding to these gaps are strongly influenced by the interpolation and the spread weights (the w_i-s). We tested Slash/A's smoothing algorithm on dense (biggest gap in the corpus: 14 days), as well as on sparse data (smallest gap in the corpus: 18 days).[6] Figure 3 illustrates the result of the smoothing of the whole Brownings' corpus (left column) and a sparse portion of it (right column).

Both dense and sparse corpora profit from the choice of weighted moving average for the smoothing. When simple (not weighted) moving average is used, in the general case only extremes are eliminated, but the "wiggliness" of the graph is not reduced: the number of angles stays about the same even by strong smoothing, they are just arranged more closely to the mean frequency value.

With sparse corpora, it is preferable to employ linearly, not exponentially weighted moving average, as this would result in a maximum smoothing level

[6] We obtained the sparse data set by selecting 12 letters from the Brownings' corpus: the first, every fiftieth and the last letters from each of the two volumes.

past which the curve cannot be smoothed. Exponentially diminishing weights become so small towards the edges of the averaging window, that the corresponding values have practically no effect on the final smoothed value. The linearly moving average technique also has a maximum smoothing level – a straight horizontal line. The problem with having a non-straight line at the maximum smoothing level is that it is not *perceived* as maximally smoothed. Besides, some of the users might want to see the mean values for the whole period, which can rarely happen when using exponential weights, but is very closely approached by the linear moving average.[7]

We also attempted to avoid the interpolation (i.e. to eliminate the spread weights in order to have $w_i = 0$, whenever $T_i = 0$). This leads to multiplication of the angles for low values of the smoothing parameter p and for sparse data sets, because the "light" missing data points tend to take the value of the closest data point creating straight line segments, connected with sharp angles.

We have chosen to take $\sqrt{T_i}$ as the value of w_i for data points and not for example T_i directly, based on the intuition (supported by the Zipf's law) that the informativity of a text does not grow linearly with its length. Alternatively, one could employ logarithm instead of square root - we have tried using logarithms with bases 2 and 10, and the results in both cases were quite similar to the ones with square root. For the missing points, the spreading of weights is best to be polynomial, as described in (2). We have tried to employ exponentially decreasing weights, but their effect is very similar to the one obtained by omitting the interpolation.

6 Visualization

We use a simple graph as the basis of our visualization since it has proved to be a powerful tool for showing tendencies over time and it does not distract the user from focusing on the data and exploring the patterns hidden in it. There are two aspects of data that are visualized on the graph – the data points that correspond to the actual frequencies of the selected n-grams and the line that represents the selected level of smoothing described in the previous section (see Fig. 3). The dots can be hidden or shown at any stage to provide the access to the original data that will be discussed below in more detail.

The information about the amount of data is displayed as a background gradient-like set of vertical lines. Each line represents one day, and the darker it is, the bigger is the number of words − to be more precise, tokens − written by the author(s) on this day. The background is author-sensitive, i.e. when the user only wants to compare the use of different n-grams by the same author, the background lines will refer to the letters written by this particular author.

[7] The maximum smoothing level for the linear moving average *approaches*, but it is not *identical* to the arithmetic average of all data points, especially in the case of sparse data, as here the interpolated values (for missing data points) are many.

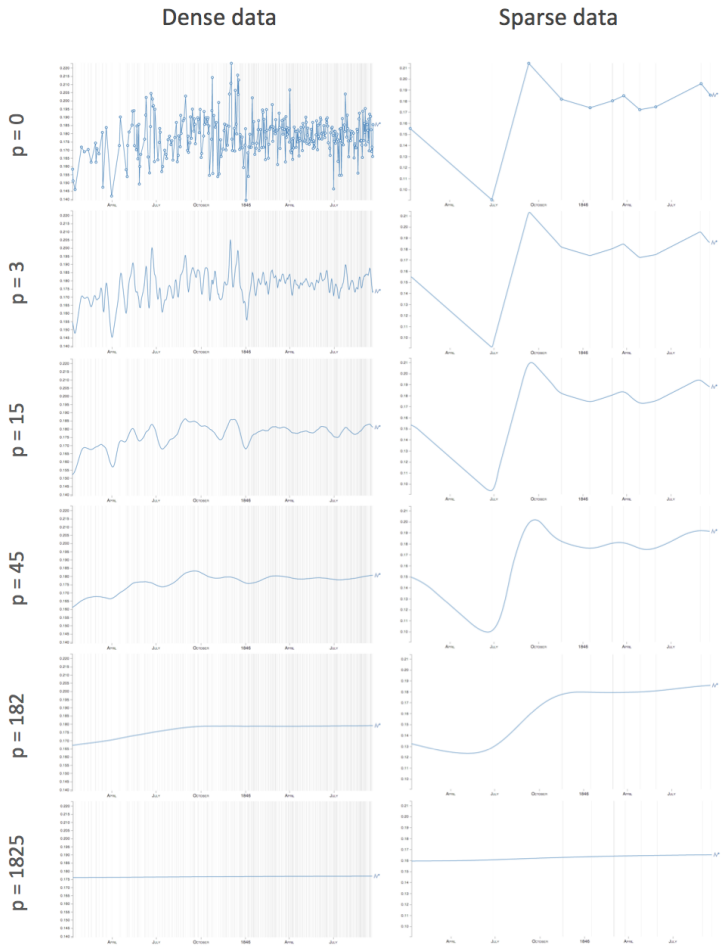


Figure 3: Comparison of smoothing of dense and sparse data with different smoothing parameters p. (p = 0 corresponds to no smoothing, i.e. actual frequencies)

Following Schneiderman's ([**?**]) taxonomy, we distinguish between several tasks that determine the functionality of the interface – overview, zoom, filter, details-on-demand and history.

In Slash/A, the graph that the user sees after loading the corpus and typing in the n-grams is an overview of the frequencies of the selected words in the specified time period. The smoothing algorithm plays the role of abstracting, or zooming out, from the original data in order to see the tendencies in the usage of a certain n-gram over time. There are six levels of smoothing in Slash/A ranging from a ragged line representing the actual frequencies (*Tendency by day*) to the last level of smoothing that shows an almost straight line (*Tendency by decade*).

If the time period covered by the corpus is bigger than a decade or if the user wants to explore another level of smoothing, e.g. tendency by two months, they can specify their own parameter following the guidelines under the *Customize* button (see Fig. 4).

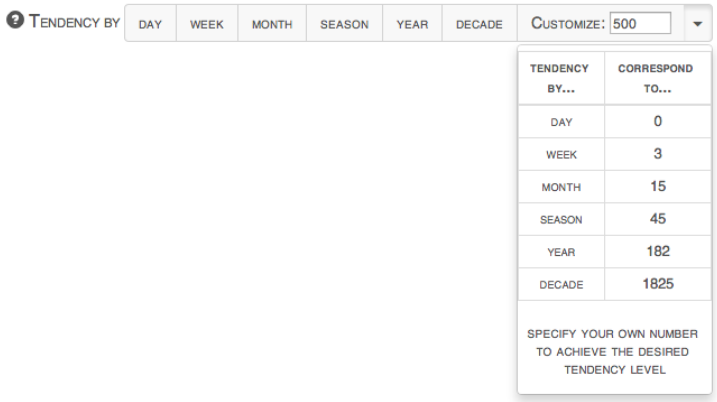


Figure 4: Several levels of smoothing represented as a tendency by certain time periods.

While analyzing the tendencies, the user might want to get access to the original text of the letters. The *Context* box gets updated every time the user clicks on any data point on the graph. It shows the metadata and the text of every letter written on the selected day (see Fig. 5).

There are also two levels of filtering that one can make use of in Slash/A. The first, initial one allows the user to only look for n-grams in the letters written by a particular author. The second option of filtering occurs with the functionality of removing the word line by clicking on the word label at the end of the line.

To make it possible for the user to trace back their queries, we introduced the *Last Queries* list at the bottom of the page. The complete history is given under the *Successful* tab. The n-grams that were not found in the corpus are listed under the *Not Found* tab. Under the *Just removed* tab one can find the recently removed graphs. Each of them can be restored with a click.

7 Input format

Slash/A is designed to process corpora in TCF XML format.[8] However, it is not necessary for the input corpus to be in exactly this format, as we only take into account certain parts of the structure of the document. We developed a set of rules that should be followed when creating or transforming a corpus to be usable as input for Slash/A. The rules are the following:

[8] A detailed description of the format can be found here: `http://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/The_TCF_Format`.

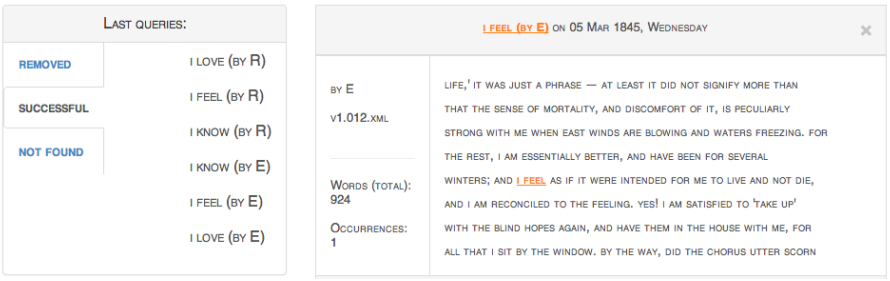


Figure 5: The *Last Queries* and *Context* boxes demonstrating the current session.

(1) The corpus should consist of XML files, each of which contains exactly one letter.
(2) The corpus should contain at least two letters written on different days. Slash/A is designed to visualize change in language over time periods longer than a day. It is not necessary for the corpus to contain at least two letters by each author that are written on different days, but if this is not the case for an author no tendency could be shown for this author.
(3) Each file in the corpus should contain exactly one node (at any place in the tree structure of the document) with tag name *correspondence* and property *from*. This property specifies the author of the letter and it needs to have a value.
(4) Each file in the corpus should contain exactly one node (at any place in the tree structure of the document) with tag name *written* and property *date*. This property specifies the date on which the letter was written and it needs to have a value.
(5) Each file in the corpus should contain exactly one node (at any place in the tree structure of the document) with tag name *text*. The data string of this node should be the letter as plain text.
(6) Each file in the corpus should contain as many nodes with tag name *token*, as many nodes with the tag name *lemma* and as many nodes with the tag name *tag*, as there are tokens in the text of the letter. These nodes can be placed anywhere in the tree structure of the document, but they must appear in the order in which the tokens they relate to appear in the text of the letter. The data strings of this nodes should be tokens, lemmas or POS tags as plain text.

Additional elements in the tree structure of the documents would neither be needed, nor have negative effect on the performance of the tool.

8 Technical notes

Slash/A is written in JavaScript and makes use of the visualization library D3.[9]

[9] `http://d3js.org/`

We have tested its performance on a corpus of 573 letters, written by two different authors in the period between January 10, 1945 and September 19, 1946.

About 12 seconds are needed for the loading of the 573 files, 16 of which are automatically excluded for inappropriate format. The processing of every single n-gram search in the rest of the files takes about 6 seconds with Mozilla Firefox 26.0 (cache limited to 350 MB) running under Linux on a CPU with 3.8GB of Ram, Intel Pentium 2020M @ 2.40GHz.

9 Future work

As we have mentioned, we used the Brownings' corpus as a development corpus for our tool. However, the generalization of the tool is only a matter of the interface adaptation since the processing of the input data is completely generalized. It will allow the user to upload their text files in the required format and make use of the available functionality of Slash/A. Apart from that, we also plan to let the user choose between case-sensitive or case-insensitive modes. We also think about introducing the option to search by recipient in addition to the already functioning searching by author. Allowing for even more precise queries (like for example specifying the n-gram's position in the sentence) would also add a lot to the functionality of the tool. In the future Slash/A can be adapted to process not only letters, but also e-mail, newspaper articles or diaries.

References

1. Carpendale, S.: Considering Visual Variables as a Basis for Information Visualisation. Research report 2001-693-16, Department of Computer Science, University of Calgary, Calgary, Canada (2003)
2. Michel, J.-B., Shen, Y. K., Aiden, A. P.,Veres, A., Gray, M. K., Brockman, W., The Google Books Team, Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A., Aiden, E. L. : Quantitative Analysis of Culture Using Millions of Digitized Books. Science vol. 331 no. 6014, 176–182 (2011)
3. Shneiderman, B.: The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. Proceedings of IEEE Visual Languages: 336–343 (1996)
4. Ware, C.: Visual Thinking for Design. Burlington, MA, Morgan Kaufmann (2008)
5. Ware, C.: Information Visualization: Perception for Design. San Francisco CA, Morgan Kaufmann (2004)
6. Yi, J. S., Kang, Y. A., Stasko, J. T., Jacko, J. A.: Toward a Deeper Understanding of the Role of Interaction in Information Visualization. IEEE Transactions on Visualization and Computer Graphics (InfoVis '07). 13(6): 1224–1231 (2007)